Developing open source tools and support materials to measure geographical accessibility to screening and cancer support services

Richard Christopher Williams
University of South Wales

Doctor of Philosophy

February 2021

Acknowledgements

Throughout this PhD there have been a lot of people who have helped. I would like to sincerely thank my PhD supervisors Prof Gary Higgs and Dr Mitch Langford for their assistance and patience. Everyone at Tenovus Cancer Care who assisted this project specifically Tim Banks and Rhian Edwards as well as the KESS team who have assisted with compassion and patience. Knowledge Economy Skills Scholarships (KESS) is a pan-Wales higher-level skills initiative led by Bangor University on behalf of the HE sector in Wales. It is part funded by the Welsh Government's European Social Fund (ESF) program for East Wales.

I would also like to thank Zoe Williams for her constant support during this time, without her it would have been impossible to undertake this PhD.

Abstract

Computing software tools such as Geographical Information Systems (GIS) have advanced the ability to measure accessibility to healthcare facilities. There are multiple methods which have been proposed in academic institutions but are often not applied in practice. The aim of this PhD is to create free and open-source (FOSS) accessibility tools which can be utilised by non-experts in the field and provide meaningful insights into the provision of cancer services in Wales. The objectives of the project are to investigate the use of accessibility methods within healthcare with a particular focus on chemotherapy and cancer services, to provide real insight into current levels of accessibility to cancer services, to use case studies to propose alternative locations and to perform user studies to provide a tool which is proven to be fit for purpose.

Multiple strategies have been used to gain access to data sets which have been used as supply, demand and network characteristics. Software solutions were investigated to establish an understanding of the current FOSS GIS available. A case study showed potential locations for additional chemotherapy resources. Modelling scenarios were implemented to investigate the effects of scale on the adopted solution in order to visualise the impacts on patterns of accessibility to inform a cost:benefit decision on planning cancer services.

Results showed multiple routes for implementation with strengths and limitations associated with each potential solution. Case studies showed that there were multiple locations which could be proposed to improve access to chemotherapy, and that a small amount of resource in key areas could have a large impact on the accessibility scores of the region. A user study was conducted with members of the Tenovus team, and this provided a real insight into how a non-expert may use these types of systems, and assisted in creating user-friendly tools to implement these models.

The study presents a software solution which offers non-GIS experts advanced accessibility methods with which to plan the location of cancer services. The system developed has the potential to change the way in which treatment is provided by allowing non-experts to calculate and visualise current and potential accessibility maps using several advanced methods. The project has investigated the impact of changes in scale, precision (network data, algorithm) and methods (variants based of floating catchment area (FCA) methods) using FOSS GIS and their networking capabilities. This tool has the potential to be used by many organisations who need to plan the provision of services in relation to changes in potential demand.

Contents

List of tables

List of figures

Chapter 1 Introduction

The body of work reported in this thesis concerns the design, development, implementation and testing of a software system that allows for the geographical accessibility of a service, specifically cancer screening and support services, to be measured and reported. One of the defining characteristics of this system will be its reliance on the use of open-source support libraries and the desire to ensure that the entire system is eventually to be made available in a free and open-source manner, not least to reflect the ethos of one of its primary sponsors, the Tenovus Cancer Care charity.

1.1 Rationale and Principal Themes

Healthcare accessibility is a well-researched theme in academic studies with a wide range of methods proposed in the health geography literature to measure access to a wide range of healthcare facilities. Findings from previous research suggest, however, that for a variety of reasons many of the more advanced methods described in such studies are not being pursued by healthcare providers and local authorities in their operational or strategic activities (Apparicio et al., 2008). The reasons for such a discrepancy between the application of such models in the academic literature and their use in addressing applied healthcare tasks are described in more detail in Chapter Two, but include the expense of proprietary software and the potential lack of computational and analytical skills in implementing such tools within some healthcare organisations. In order to address these gaps, this project aims to provide a free and open source software solution which can perform advanced accessibility methods and be used by non-experts within healthcare organisations to address real world problems. Whilst the particular application area to hand is the measurement of access to cancer services in Wales, these tools have been purposely developed to be transferable to other types of (healthcare) application areas in a wide range of contexts.

The Open-Source Initiative describes open-source software as: “software that can be freely accessed, used, changed, and shared (in modified or unmodified form) by anyone”. Open source software has been written and developed by a wide range of individuals in a variety of interest groups, and is distributed under licenses that comply with the Open Source Definition[1].. Free and Open Source Software (FOSS) such as, for example, Linux (operating system), Mozilla Firefox (browser) and VLC Media player, has been shown to be beneficial to a wide

[1] https://opensource.org/faq#osd

range of organisations, with those such as Amazon, Google and Salesforce using such solutions (Fitzgerald, 2006a). In particular, having a shareable, customisable program which is free of cost can benefit potential users who do not have the budget for proprietary software, or would prefer to use those types of resources elsewhere. There are many examples of FOSS geographical information systems (GIS) which are being used by a variety of organisations to address real-world applications. Software such as GRASS GIS, QGIS and gvSIG for example have been used to address a wide diversity of mapping and analytical tasks (see section 2.5 and 2.6).

Since the 'quantitative revolution' that took place within academic Geography in the late 1960s, driven by advancements in computing power and its wider availability in universities and academic institutions, considerable theoretical development and a great many practical methodologies concerning the measurement of geographical access to services have been proposed. Many of these are discussed in greater detail in Section 2.4. Amongst these, a relatively recent methodological development is a set of techniques that may be broadly grouped under the banner of Floating Catchment Area methods (FCA), and it is these that feature particularly strongly within this body of work. Overall, they are essentially all variants of a broader methodological approach, namely "geographical gravity modelling" which again had its roots in the 'quantitative revolution' of the 1960s and 1970s.

With regard to the implementation of these models, Langford et al[2] developed an advanced accessibility tool using the ArcGIS proprietary platform. This software was used in a wide range of application areas (Higgs et al., 2018; Langford, Higgs and Radcliffe, 2018) and offers an advanced accessibility tool to those who have access to ArcGIS software. Working in conjunction with Tenovus Cancer Care, this project aimed to build on such research in order to develop open-source solutions offering similar functionality, implementing FCA tools using FOSS and open data sources. To realise the maximum potential benefit of such methods, it was important that such programs were designed to be as easy to use as possible by non-experts who are charged with planning the provision of such facilities, by keeping many of the complexities hidden, limiting the number of inputs and managing the data. Usability studies were first introduced in the 1980s and have become an important part of software development cycles. By accessing potential users of such tools within Tenovus there was an opportunity to

[2] https://www.researchgate.net/publication/287198887_USW-FCA2_An_ArcGIS_add-In_tool_to_compute_Enhanced_Two-Step_Floating_Catchment_Area_accessibility_scores

develop software in conjunction with healthcare professionals that could be used to aid the planning and management of cancer facilities. By exploring their needs and running usability studies in collaboration with these actors at key stages in the software development cycle, this work sought to enhance the ease of use of the proposed system, the intuitiveness of the interface, and the appropriateness of the manner in which outputs are presented.

This research investigated the provision of healthcare services within Wales with a particular focus on current and projected levels of accessibility to chemotherapy services. This study provided an overview of current levels of accessibility and investigated the potential for improving accessibility to provide services as local as possible to cancer patients. Using open data, and that gathered routinely and collated by Tenovus, the over-arching aim was to investigate current accessibility levels across Wales and draw on case studies that investigated ways of increasing accessibility to such healthcare facilities. This project aimed to create open-source software that allows users to perform complex healthcare accessibility calculations based on floating catchment area approaches. There is a real need for the application of such tools to reap the benefit of advanced accessibility approaches, and provide healthcare organisations with the types of models and associated expertise to plan the provision of healthcare services. Furthermore, making this code more widely available enables academics and policy makers alike to have access to the source code to not only to run the initial program, but also to then customise it to their specific needs. Models of this nature not only help users to map the current provision of healthcare in their area of interest but can also demonstrate the potential effect of increases or decreases in capacity on access to services within the surrounding areas and help in planning wider healthcare provision within local communities.

1.2 Research Aims and Objectives

The overarching research aim is:

To develop free and open-source software which can be used by non-GIS experts to model accessibility to healthcare (chemotherapy) services and provide meaningful insights into the provision of cancer services in Wales.

To address this aim, the following supporting objectives were explored during the course of the research:

- To examine the utility of accessibility models based around the use of floating catchment area (FCA) methods in healthcare studies with specific emphasis on the provision of cancer services.
- To investigate and evaluate the development of solutions based on the free and open-source ethos, that can be utilised to help administrators and managers and other non-GIS experts to plan the provision of healthcare (cancer) services.
- To illustrate the practical application of these tools by using data on the current availability of cancer care services in Wales, using implementations based on FCA solutions.
- To collate and analyse a number of datasets which can be used to give a meaningful insight into the current provision of chemotherapy services as they exist within Wales.
- To demonstrate the use of GIS-based accessibility models to examine the supply of mobile services in relation to potential demand for cancer services.
- To develop user-friendly software to enable non-specialists to explore the implications of alternative models of provision of mobile services.
- To test the usability of these tools with professionals concerned with helping to deliver cancer services.
- To draw on these findings to show how such programs can be used by the profession to plan and manage the provision of a wider range of healthcare services.

1.3 The provision of cancer services in Wales

Macmillan estimates that there are 160,000 people living with cancer in Wales[3] and this figure is projected to grow to 220,000 by 2030. Several reports have been produced by NHS Wales and different specialist charities on the current performance and expected improvements of healthcare services within Wales for the treatment and prevention of cancer, and these have highlighted several areas of focus. The Cancer Delivery plan for Wales, (*Cancer Delivery Plan for Wales 2016-2020*, NHS Wales 2016) highlighted several key actions for service delivery that have direct relevance to the research conducted here:

Key action 21 (page 12) "The regional cancer centres, Wales Cancer Network and Welsh Health Specialised Services Committee (WHSSC) to work together to ensure equity of access and delivery of service quality".

[3] https://www.macmillan.org.uk/_images/cancer-statistics-factsheet_tcm9-260514.pdf

Key action 31 (page 12) “Healthcare boards to continue to monitor performance against national waiting times targets and focus on improving performance”.

Key action 32 (page 13) “Healthcare boards to ensure cross border flows do not disadvantage patients in access to treatment and care.”

Key action 40 (page 15) “Where possible healthcare boards should provide care locally and support patients who need assistance to travel or stay away from home”.

This project has particular interest in key actions 21 and 40. The need for NHS Wales to be able to plan service provision is of great importance to them as well as being able to understand the equity of service across the different healthcare boards that operate within Wales (figure 4-11 displays the areas of each healthcare board). A system that is able to assist in both the planning and evaluation of service provision will be of great help when improving services and monitoring outcomes from the cancer delivery plan

The Cancer Delivery Plan for Wales highlights the need for further research and for more of this research to be brought into practice. For example, it suggests a need for “Assisting organisations in the planning and delivery of their services, ensuring that research results translate into evidence-based practice” (page 18). Thus, the need for research to be utilised and put into practice to improve current service provision in Wales is highlighted as a key driver for achieving the goals set out in the Plan.

Cancer Research UK has called on the Welsh Government to “ensure that the new organisational structure for cancer services provides clear leadership and accountability mechanisms” p.4. They further recommend that “(T)he Welsh Government should review the existing approach to commissioning specialist treatments, such as radiotherapy, chemotherapy and low-volume surgery. The new strategy should establish a national commissioning body to better plan and coordinate these services across Wales. The new strategy should also set a clear ambition to improve access to clinical trials across Wales, and detail of how this is supported.” Furthermore, they suggest that “(T)he new strategy should develop a national dataset for chemotherapy and radiotherapy activity. LHBs will need to supply the information and data completeness should be reported via the annual national cancer report."[4] The work that Cancer Research UK is calling for would assist NHS Wales in their goals, and highlights the need for there to be central co-ordination between healthcare boards regarding chemotherapy and other treatments. Having a centralised source of data would assist in the ability to co-ordinate across regions which is of particular importance when some healthcare boards have limited chemotherapy provision (discussed further in section 4.3.2).

The Wales Cancer Alliance produced a report Priority Policy Calls[5], which highlighted many areas for improvement. The Alliance is made up of 20 different charities that operate in the region and the report hoped to act as a guide to future cancer policy. The report highlighted areas for improvement in prevention, detection, and treatment. It called for action to tackle inequity in access (page 6) and for infrastructural support for analytical platforms (page 8). These among other policy suggestions were included to help steer the quality of future cancer care and improve outcomes.

The Wales Cancer Alliance echoed the calls made from Cancer Research UK and the actions set out in the NHS Wales Cancer Delivery Plan. Equity of service was emphasised in their report, and the need to measure this spatially highlighted the difficulties in monitoring such activities with the current data available. Understanding supply and demand in each healthcare board, and the effect that this has on the whole country, is seen as the key to achieving the goals set out.

The elements that have been put forward within the various documents addressing cancer care in Wales are wide ranging, but key for this project is that collaboration between organisations is required to plan service provision. This, alongside the need for treatment to be closer to home provided a strong basis for this project to assist and address some of the key challenges that are facing cancer treatment provision in Wales.

1.3.1 The role of Tenovus Cancer Care in the provision of services

Tenovus Cancer Care (Tenovus) is a charity which was started in the 1940s by ten businessmen. The charity focused on the South Wales area and funded a variety of projects. In the 1960s they formed the Tenovus Institute for Cancer Research in Cardiff, and this narrowed

[4] https://www.cancerresearchuk.org/sites/default/files/cancer_services_in_wales_full_report.pdf
[5] https://walescanceralliance.org/wp-content/uploads/2019/05/Overview-Document.pdf

the focus of the charity to concentrate specifically on cancer related issues. The centre has since created two cancer drugs which are now used worldwide to treat breast and prostate cancer. In 2018/19 Tenovus managed to fund over £500,000 of cancer research and supported over 6,500 people affected by cancer. In this period, they raised over £9.5m. Tenovus Cancer Care have 5 strategic aims[6]

- To provide support and treatment to people with cancer and their loved ones, closer to home in unique ways.
- To represent the needs of people with cancer and their loved ones.
- To conduct and fund research to improve cancer outcomes and experiences.
- To work with communities to develop healthy lifestyle programs, raising awareness of how to prevent cancer.
- To develop a committed and engaged team that manage their resources effectively to support our work.

In 2009 the charity launched a mobile support unit which provides support to cancer patients and their families. The charity currently has two mobile units, one delivers chemotherapy services at multiple locations around Wales, and the second concentrates on lymphoedema treatment and education. A third unit was in circulation to provide prostate cancer services in Wales, although no treatment was available on it, and it has been removed from circulation.

In England and Wales Tenovus operates a number of choirs for cancer patients and their families. They have conducted research into the positive effects that singing can have and were awarded a £1m grant to set up further choirs. They also selected a number of cancer victims to be part of a BBC documentary that followed the members of the choir from the initial start to a performance at the Royal Albert Hall. Tenovus now operates a number of services for cancer patients and their families. These include a cancer call back service, where an experienced nurse will call the patient at regular intervals to make sure that they are supported throughout the entire process. They offer a counselling service which can be used face to face or over the telephone, as well as a welfare advice service which can be accessed by people for social and financial support. Tenovus operates several outreach projects at different points throughout the year. These range from a converted ice cream van which raises awareness of skin cancer all

[6] https://www.tenovuscancercare.org.uk/about/about-us/

over Wales, to having a presence at Greenman festival to promote cancer awareness, and a number of other appearances. In addition, they have a number of shops in England and Wales which provide fundraising and brand awareness. They also employ a number of inventive ways to raise funds, from charity runs and cycle rides, to bungee jumping and sky diving. They have a large database of donors and actively contact their supporters to continue to raise funds. The funds they raise are split between raising cancer awareness, operating the mobile units and other support services; and the funding of a number of research projects.

In a policy context, Tenovus have used their position within Wales to draw attention to issues surrounding the provision of services to the National Assembly with their manifesto "More than medicine"[7]. This highlighted many key areas such as support for those with cancer and on-going support for families in the event of their loved-ones passing away. The manifesto suggested that prevention is better than cure, and it tackled healthcare issues related to the use of tobacco, fast food and alcohol, as well as some other areas of concern. One of the priorities, as highlighted in The Cancer Delivery plan for Wales, 2016-2020 (NHS Wales), was the need to provide treatment closer to home utilising their mobile facilities.

1.3.2 Tenovus mobile units

Tenovus has two mobile units which deliver cancer care in the community at a number of sites in Wales (section 4.3.2). These sites were chosen in conjunction with the healthcare boards that supply the nursing staff, so the units are currently not spread evenly throughout Wales, and most of the work is conducted in South Wales. The locations are normally areas with good access; in Cardiff for example they use a large supermarket carpark so that the cancer patient can park near the unit and make the process as painless as possible. Each unit is able to provide a maximum of 54 treatments a day. This is because of the size of the unit and the number of people they can facilitate whilst still providing the level of service needed. Each day the chemotherapy unit works on a specific type of cancer so that specialists can be used. They are dependent on the healthcare boards for referrals; so being able to show where they are able to make the most difference to the healthcare boards is important from an overall policy perspective. The availability of the latest mobile unit added to their overall provision of services

[7] https://www.tenovuscancercare.org.uk/wp-content/uploads/2020/06/tenovus-cancer-care-more-than-medicine.pdf

has been covered extensively in the press and is one of the largest mobile chemotherapy units in the world[8].

1.3.3 Spatial Mapping and Analytical capabilities within Tenovus

Tenovus had no mapping capabilities at all. They were guided by the healthcare boards and the locations that were available when choosing the sites for the mobile units and for the other services they provide. This was aided by the knowledge of Tenovus staff, in particular their familiarity with potential locations for siting such facilities across Wales. Data was manually inputted into two different databases which could then be manipulated to provide an overview of the services they provided. This data capture was not very reliable and the recall of data was difficult for them.

After meeting with key stakeholders, it was clear that Tenovus operated a very secure system due to the sensitive nature of the data that they held. Initial discussions regarding the implementation of open source programs however suggested that the IT section of Tenovus did not foresee any significant problem with the adoption of open-source software as long as it was operating fully within the corporate firewall.

1.4 Thesis Chapter Outline

Having introduced the main topic of this research and outlined some of the broader context in which it took place, the remaining thesis is structured using the following chapters:

Chapter Two: Describes the relevant literature for this study; a brief history of accessibility studies, the ways in which FCA methods are utilised and provides an understanding of the different enhancements to initial formulations of the FCA methods approach. This is key to creating a useable tool which can provide insights into the current provision of chemotherapy in Wales. To do this a review of historical healthcare accessibility methods was undertaken prior to an in-depth investigation into the implementation and application of the 2-step floating catchment area (2SFCA) method. Developments leading to the use of enhanced 2 step floating catchment area (E2SFCA) methods are then described and their use in the healthcare arena are discussed. Previous studies that have used such approaches to examine access to cancer services are described in order to place the current research in the context of those studies that have used FCA approaches to analyse different facets of cancer care. The chapter highlights

[8] https://www.bbc.co.uk/news/uk-wales-45969165

the use of different FOSS within healthcare research and summarises how GIS approaches have been used to date within studies related to the provision of cancer services.

Chapter Three: A standalone chapter discussing the development of the proof-of-concept tool which occurred using QGIS. It covers the initial development and testing, and allows for an investigation into how an accessibility tool could work; and create a pathway for the standalone version. This chapter discusses the results gathered and critically examines the shortfalls of this implementation, providing an overview of lessons learnt and how this can be taken forward to the next stage of development. Initially an overview of QGIS and the development environment is discussed. To develop plugins in QGIS, Python and PyQt are used. The benefits and problems of using Python for development are discussed as is PyQt. Straight line 2SFCA is discussed as this solution does not have the capacity for a routed network solution. The data used at this stage of the project is discussed and highlights any weaknesses found. The outputs from the tool including maps developed for Tenovus are highlighted and then a description of the performance of the tool is conducted. Various tests were completed to understand the capabilities of the tool, including the use and application of very large data sets in order to understand the effect of looping through the data and the need for efficient organisation of the database. Finally, a discussion around the strengths and weaknesses was conducted and the key findings from this system were described to help guide the planning of the main tool. Several of the available choices for development are discussed and different options are presented which are explored in more detail in the methods section.

Chapter Four: The methods section follows the proof-of-concept section and investigates those FOSS options available for further research and development. As many solutions as practicable were investigated and the main pros and cons were documented. A solution to move forward is presented and the rationale behind the chosen approach is discussed in more detail. There are multiple potential solutions to address the aims highlighted in this project and the rationale for the approach adopted in this study is fully documented in this chapter. As with the FOSS GIS solution there were many ways in which the routing could work and the different choices are described, as are the methods used to evaluate them. As with the routing choices there were a number of different ways in which the algorithm could be implemented and a discussion surrounding this is conducted. A case study was completed to understand the effects of moving the Tenovus units to different areas or adding additional supply to selected locations. This was done by using a dataset of all the large supermarkets, and associated parking space suitable for siting mobile units in the UK, and allows specific locations to be investigated in the analysis. There is a large body of work surrounding the datasets that are utilised by the program. Initially, the supply side data was discussed as were the data sets received from a Freedom of Information (FOI) request. This data gave a good understanding of the current level of chemotherapy within Wales. The demand side was also investigated to understand how people utilise these services. The data ranged from census and deprivation measures such as the Welsh Index of Multiple Deprivation (WIMD), to a full dataset of the users of Tenovus' services, this allowed for a good understanding of the spatial patterns in demand for cancer services and informed the choice of parameters such as catchment size and some other key information. A further discussion around the way in which the data was visualised and how best to present the data for inspection, as well as the documentation used to ensure that it is suitable for the non-expert user is included here.

Chapter Five: There is a standalone chapter which discusses usability aspects of this project. Initially a literature review was conducted, to discuss the history of usability and the current models and methods that are used in similar studies. Understanding the current methodology, enabled user studies to draw on suitable methods and models to ensure that the user was at the centre of the design process and that this tool is capable of being used by the non-expert user. The first user study employed three main approaches to gather information which included a questionnaire, one-to-one interviews and a focus group with the participants. Findings from this stage informed subsequent developments, and these were highlighted before implementation of the second iteration of the tool. The second user study used similar methods, and provided similar feedback and discussions surrounding the results. The implementations and changes to initial versions were noted and further feedback highlighted. The results of the user study are discussed so that it is possible to track changes across the different iterations of the solution.

Chapter Six: The results section follows on from the methods section and discusses the findings of the study. Initially the results of some performance testing were described to show the times for a variety of sizes of dataset. A comparison between this tool and the Langford et al tool developed in ArcGIS was conducted to see if there were any significant differences between the two. The results of the case study are described showing the effects that adding additional Tenovus chemotherapy provisions to additional areas has upon the total accessibility for those needing chemotherapy within Wales. The results are displayed with visualisations and the differences are highlighted. Many of these visualisations show the impacts of varying data scale, using alternative road networks and the use of different network connections and FCA methods. A summary of the documentation created is also included.

Chapter Seven: In this chapter the results of the project are reviewed and discussed in more detail. This process puts the results into context and relates them back to the aims and objectives of the project as set out in Chapter One. Key methodological and policy aspects of the findings are highlighted and the wider implications are discussed in relation to findings from previous studies (reviewed in Chapter Two).

Chapter Eight: The conclusion summarises how the research aims were addressed and discusses limitations of the study and potential ways to address such concerns. There is a summary of the research findings, a reflection of the approach adopted in the study, and a re-iteration of the policy relevance of this research for Tenovus and other organisations looking to adopt such methods to plan the provision of healthcare services. This chapter also includes some recommendations for future work and draws attention to the new contributions to knowledge revealed in the study.

Chapter 2 Literature review

2.1 Introduction

Arguably, healthcare accessibility is fundamental to a population's health and wellbeing, as understanding the relationship between the healthcare supplier and the population it serves is important in healthcare planning and disease control. Historically, the methods in place to measure healthcare accessibility have varied in their accuracy and complexity such as person-provider ratios and shortest distance calculations (Guagliardo 2004, Luo 2004) which rarely give a detailed understanding of accessibility (over simplification, assumption of boundaries and Euclidean distances). The two-step floating catchment area (2SFCA) method is a special type of gravity model that has the potential for a more advanced assessment of changes in supply demand relationships mitigated by distance, and hence provides a more nuanced measure of accessibility (Luo & Wang, (2003), Luo & Qi, (2009)). Such methods are commonly applied in a range of application areas, largely as a result of advancements in geographical information systems (GIS) which allow for large amounts of data to be analysed quickly, such that system calculations that were previously impractical are now being implemented with relative ease.

This chapter firstly investigates the use of accessibility models in healthcare applications before introducing the FCA approach, and providing a timeline of changes and refinements since earlier formulations. It includes a description of gravity models, and highlights how 2SFCA is a special case of the gravity model. These are compared to the use of 'traditional' measures of accessibility within such studies. To understand the implications of the changes in method it is important to discuss the origin of these methods and to provide a brief history of their application particularly around healthcare geography. This section will also draw on findings from the use of floating catchment area methods in other thematic areas.

2.2 The use of Accessibility Modelling in Healthcare Studies

There are many ways to define healthcare accessibility, but for the purpose of this thesis the definition by (Khan, 1992) is used where he defines it as "availability of a service moderated by space, or the distance variable" (p. 275). Measuring access to healthcare is important for a number of factors, and it has been performed in numerous ways depending on the available computing systems, data availability and the application in hand. Apparicio et al., (2008) identified five commonly used measures of spatial accessibility:

- distance to closest service
- number of services within a specified distance or travel time
- mean distance to all services
- mean distance to a certain number of closest services
- gravity model-based measures

The first 4 measures above were relatively easy to compute. For instance, the distance to the closest service could involve the measurement of the distance between a population weighted centroid and the nearest service point, leading to assessment of whether the distance was within a specific threshold and is therefore deemed “accessible”. All of the above are commonly used in accessibility literature because of their simplicity to compute in a GIS environment. The increasing functionality of GIS has, however, meant that more complex methods have been proposed to measure accessibility.

Joseph and Bantock, (1982) used the gravity-based model to measure healthcare accessibility. The main difference between the gravity-based model and 2SFCA is that there is no catchment area in the gravity model, rather a continuous decay coefficient is used. This means that a supply point outside of a catchment was not counted in the 2SFCA calculation. The gravity model does account for it, but the further away the supply point is the less influence it has. In theory, this method explains accessibility better than the 2SFCA models, but it is more complicated to compute and not as intuitive to use (Luo and Qi, (2009). Luo and Wang (2003) showed that the gravity-based model tended to overestimate accessibility in low access areas, and these are commonly the areas of most interest in healthcare accessibility studies. Schuurman, Bérubé and Crooks, (2010) proposed a modified gravity model which had two changes to the original, it used travel time instead of distance and utilised a continuous inverse-power impedance function.

Other models have been used to study healthcare accessibility. Spatial interaction models are based upon the gravity model and attempt to determine the flow of people through the knowledge of the demand (origin-constrained model), supply (destination-constrained model) or both (doubly-constrained model). Cromley and McLafferty, (2002) suggest that the destination-constrained model has been used extensively in the UK due to the central nature of the NHS and government, whereas the origin-constrained model has been more widely used in the USA as there are no central institutions in control of the healthcare system. These models have been used as the basis of many more models which aim to assess accessibility throughout the years.

The field of location-allocation modelling is large and has been used to address healthcare accessibility in numerous ways. There are problems which location-allocation modelling attempts to address, Tomintz et al., (2008) investigates the best way to allocate stop smoking services within Leeds, England and uses the p-median problem that defines a number of demand sites and service allocation is optimised so that all demand points reach only one supply location, for this different data is required. The number of demand points, number of potential supply sites distance or cost (network) of travel between supply and demand sites and the number of facilities that are to be opened. The heuristic method was used to solve the problem and when the results are overlaid with the real use case it is possible to see how this method would have placed the services to reduce travel time for the demand. The authors note that this is at a very small scale and that this study used the populations home and did not account for those who had to travel for work.

Kelly et al., (2016) conducted a review of papers regarding distance to healthcare and healthcare outcomes in countries in the northern hemisphere. Out of the papers reviewed, 77% found a correlation between distance or time travelled and healthcare outcomes where the further the distance the poorer the healthcare outcomes. This suggested that proximity to healthcare facilities has an impact on healthcare outcomes.

These methods are not exhaustive and many different models could be used within this project. There is a large body of work on healthcare accessibility as it plays an important part in healthcare outcomes. This project uses the FCA method and the following sections elaborate on the literature surrounding this topic.

2.3 The use of GIS approaches to investigate variations in access to cancer services

There is a wide literature on the use of GIS approaches to study variations in access to cancer services and spatial patterns in cancer diagnosis, treatment and survival. The findings of such studies provided important context for this study which focused on the use of FCA approaches by highlighting different approaches to investigating access to cancer services.

Rigby and Gatrell, (2000) used breast cancer data from 1982-92 from the North West Cancer Registry. The data provided the location at diagnosis of each patient within North West Lancashire and was used to perform several exploratory analyses. This information is personal and very difficult to obtain, and exploring clustering effects with it provides insight into an area which is difficult to study. 3 scales of analysis were completed, first at local authority ward level, then at Enumeration District level (around 130 households) and finally at individual point level. The results showed that there were some areas of cluster and that a potential rural/urban split occurred within the first two scales. At point level it was possible using 3D visualisations to pinpoint several areas with high clusters. This paper highlighted the potential of non-aggregated patient data, and showed it is possible to gain insights into spatial patterns with data at a finer scale.

A study investigating the effects of socioeconomic status on colon cancer survival was completed which compared Toronto, Canada and San Francisco, USA (Gorey et al., 2011). The study analysed registry data from locations of people diagnosed between 1996 and 2000, and followed the outcomes until 2006. The study highlighted the fact that income was directly associated with many stages of treatment and survival in San Francisco, but did not have the same effect on the Toronto data set. This study showed that there are metrics of interest which can be used to assist in evaluating cancer services success. Evaluating socioeconomic variance across locations gives a richer understanding of the data, and can assist in service provision planning.

Lyimo and Beran, (2012) conducted a study in Tanzania to establish the factors associated with cervical screening. 354 women were interviewed for this study and the researchers asked numerous questions in questionnaire format at the participants' home. The study found that 22% of women had been screened for cervical cancer even though the prevalence is high in the region; the study noted that only access to screening and the knowledge of the cancer and treatment were significantly associated with screening uptake. The authors highlighted several limitations with the study; the region chosen for the study may not be generalizable, and some of the questions may have seemed leading. This study highlighted the importance of access to cancer services and showed that visibility of treatment is key in screening uptake.

Furthermore, a study concerned with rural-urban disparities in late-stage cancer risks was conducted (McLafferty and Wang, 2009) in Illinois, USA. Multilevel modelling was used and the study looked at the major types of cancer in breast, lung, prostate and colorectal. The paper highlighted that the levels of stage 4 cancer were highest in urban areas, with the most rural locations also having an increase in rates. The study showed the need for there to be higher levels of screening and education in urban areas.

Adams et al., (2015) conducted a large-scale study investigating the link between accessibility to health centres and cancer mortality to incidence ratios. An analysis of breast, colorectal, cervical and prostate cancer at county level in the USA between 2006-10. There are 7,240 federally qualified health centre delivery sites across the nation, and these were used to measure access. The study showed that access to healthcare improved the mortality to incidence ratios across the types of cancer, and noted that the federally qualified health centres disproportionately served low income and ethnic minority populations. The authors suggested that the paper was limited in its reach and should be used to assist in hypothesis generation, as opposed to being a decisive piece of work, as there were multiple variables that the study was unable to include. Nonetheless, the paper showed that there is link between cancer survival rates and access to healthcare facilities, and added weight to the argument that it is good practice to provide healthcare services as locally as feasible.

Similarly, to Adams et al., (2015), Walters et al., (2013) conducted a large-scale analysis of breast cancer survival across six high income countries: Australia, Canada, Denmark, Norway, Sweden and the UK. The study analysed data on over 250,000 women across the regions and highlighted a difference between the UK and Denmark when compared to the other 4 regions, as they had a lower survival rate. The authors suggested that there were two reasons for a higher mortality rate, and that lower survival rates occurred when either there was an advanced stage of diagnosis, or stage specific survival was low. The paper proposed more investigation into international differences to identify areas for improvement. This study highlighted the need to evaluate and compare data from different regions using different methods and that these comparisons could highlight areas for improvement.

On a smaller scale, Belasco et al., (2014) conducted a study on the impact of rural healthcare accessibility in relation to cancer related behaviours and outcomes. The study focused on the high plains region of Northern USA and used data for 2005-09. Principal component analysis was used to establish a healthcare accessibility index, and a regression was performed. The results showed that the areas in the study with poor access had a lower breast screening uptake, higher risk-related behaviours (smoking, obesity) and higher rates of cancer related deaths. The authors suggested that there are two focuses, increase cancer related prevention activities and reduce risk related behaviours. This paper showed the need for good accessibility to cancer services, and showed that poor behaviours were more prevalent in areas with poor accessibility.

Khan-Gates et al., (2015) completed a systematic review of literature relating to breast cancer screening and accessibility. The paper found that there were many variations in the studies completed but that higher accessibility to mammography resulted in lower stage diagnosis. The authors suggested that more work needed to be done with a particular focus on actual travel patterns and utilised care. This study highlighted that there is much more work required in this area to understand the effects that accessibility has on cancer survival rates.

Floating Catchment Area (FCA) Methods: An introduction

The 2SFCA method builds on the previous implementation of the floating catchment area (FCA) method. Peng (1997) used a catchment area to float over each population point in a study investigating job accessibility. This study used a square catchment area and was later changed to a circle by Wang, (2000). Whilst this method improved the standard supply to demand ratio by moving the catchment over the study area, it only accounted for one side of the supply and demand so it was unable to provide a clear picture of accessibility. Radke and Mu, (2000) first introduced the 2SFCA model to measure access to social services, and in this study Euclidean distance was used to compute the catchment sizes. The use of a routable road network has become common practice as this gives a richer understanding of the travel times within areas.

The initial formulation was presented in a study by Luo and Wang (2003) based on a paper by Radke and Mu (2000). The following section of the chapter describes the initial formulation of the approach before describing how it has been used in previous studies. Following the discussion of 2SFCA, a detailed evaluation of the enhanced two-step floating catchment area (E2SFCA) is described and the areas that it was intended to address are evaluated to highlight the strengths and weaknesses of the enhanced method, drawing on the work of Luo and Qi (2009).

2.4 Implementing Two Step Floating Catchment Area (2SFCA) Models

This section details papers from the initial formulation of the approach; including the methods used, thematic area, and whether the approach adopted a spatial approach or included the use of aspatial data. The list of papers is not exhaustive but covers a period of rapid expansion in 2SFCA methods which claimed advancements to the original 2SFCA approach. Most of the papers on 2SFCA have been describing a similar method on different data in different regions with only minor changes.

2SFCA is an advanced accessibility method which in simple terms, accounts for both proximity to a service and the amount of competition for those services. The first step defines a catchment area around each supply point and sums the contained population to get a supply to population ratio. The second step defines a catchment area around each demand point and sums all contained supply points from step 1. This score can then be visualised using a GIS to provide an insight into accessibility which accounts for both distance and demand for the service.

The two-step floating catchment area method is a special case of the gravity model and was first introduced by Wang, (2000). It has been used in several studies of healthcare accessibility since its original application in this field by Luo and Wang, (2003) who were the first to compare the results of 2SFCA to those of a standard gravity model. In summary, the 2SFCA process uses two 'steps' to calculate an accessibility score. Step 1 sets a default travel time (or distance) from a (healthcare) provider, and sums up the population that can access that provider within the specified travel time to create a provider to population ratio. Step 2 uses a population centre and sums up the ratios to provide an accessibility index. One of the main advantages of this method is that it is easy to interpret and accounts for the interaction between supply and demand. Luo and Wang (2003) highlighted the assumptions in this model (such as uniform accessibility anywhere within the threshold catchment – there were no distance decay parameters in the original formulation). Nevertheless, their study did identify advantages of floating catchment area techniques over 'traditional' measures of accessibility. Luo and Wang (2003) describe the 2SFCA method as follows:

Step 1: Calculate the provider to population ratio Rj at each location j:

$$R_j = \frac{S_j}{\sum_{k \in \{d_{kj} \leqslant d_0\}} P_k}, \quad (1)$$

Step 2: Calculate the spatial accessibility index Afi for each population site i:

$$A_i^F = \sum_{j \in \{d_{ij} \leqslant d_0\}} R_j = \sum_{j \in \{d_{ij} \leqslant d_0\}} \frac{S_j}{\sum_{k \in \{d_{kj} \leqslant d_0\}} P_k}, \quad (2)$$

Notation key

Sj: medical capacity at each provider j

Rj: provider-to-population ratio

Pk: population site k

do: threshold travel time

Dkj: travel time between k and j

Afi: Spatial accessibility index of each population site j

Dij: travel time between i and j

In their study, Luo and Wang (2003) compared the access scores from a gravity model approach to that of the 2SFCA approach. They used census data from the Chicago area (including a 15-mile buffer zone) and purchased data on primary healthcare physicians in the area. A GIS environment was used to set out the fastest routes possible to enable the calculations. They created a population weighted centroid and used the zip code centroid for the physicians; and found some differences between how the two different models reported their results. The main point was that the gravity model was more likely to show smoothing, this is where areas with low accessibility are missed. This was found to be more prevalent the larger the area being used. They concluded that “based on this preliminary case study, we recommend the two-step FCA method, simpler and easier to interpret, for use in improving the designation of healthcare professional shortage areas” (Luo and Wang, p. 870). They also identified the following areas for further research and highlighted some limitations of the model;

firstly, the research assumed that everyone had access to private transport. Secondly, an investigation into the variation of accessibility correlated to that of socio-economic status and the ethnicity of the population, showing that there was a pattern of poor accessibility in deprived or ethnically diverse areas. Thirdly, they highlighted the importance of investigating longitudinal changes in accessibility scores by suggesting that more research was needed to look at the differences in accessibility scores between 1990 and 2000.

As the authors mentioned, this method did have flaws that were redressed in a wide range of later studies that applied floating catchment area techniques. For example, it assumed that all centres within each catchment area were equally accessible, this was not the case as there was still a difference in accessibility within each catchment. McGrail and Humphries (2009) showed that in their study area in Victoria, Australia, for large catchment sizes the method did not account well for populations in rural areas. This method assumed boundaries that were unable to be breached and someone that was just outside of a catchment area was unaccounted for. This method also assumed that everyone had the same access to transport, and by using driving time it did not account for those that used any other means of travel.

Application of FCA approaches in Health Geography

Two step floating catchment area (2SFCA) methods have been widely used in studies that have involved examining access to healthcare facilities. This section provides a snapshot of the types of application areas that have been shown to have benefitted from the use of these types of models. Song et al., (2013) used 2SFCA to assess the availability of Maternity units in Shenzhen, China. The study area had 81 maternity units and the demand was abstracted as a percentage from the sub-district populations. The paper found there to be two areas with poor accessibility, Nanshan and a large part of Langgang. The authors recommended additional resources for these areas. This study did have some limitations, the demand data being abstracted and not being real data, the use of centroids to represent the demand, the supply data and the demand data were from different years, and everyone within the 20km travel distance was deemed to have the same level of accessibility as no distance decay functions were used. The authors highlighted the fact that although there were limitations the study had positives as it was a more robust method than the ones currently in place. The paper involved an interesting comparison between providers of maternity units and, by comparing the resource from private or state units, the authors were able to understand the effects that the private sector had on overall provision. This ability to show the different types of provision was particularly relevant to the study with Tenovus as it was key to understanding not just the overall accessibility of an area but to also evaluate the impact that access to the Tenovus mobile units was having.

Ngamini Ngui and Vanasse, (2012) used 2SFCA to investigate access to mental health provisions in South West Montreal. This small-scale study investigated accessibility to mental healthcare provisions at a relatively fine scale. The paper identified several areas where access was poor such as Dorval and Saint-Henri which had little access within 3km. The authors recognised some limitations to the study; aggregation-error was an issue which they attempted to overcome by integrating dissemination block data to better represent the spatial distribution of individuals and edge effects on the study area. The study also did not account for any distance decay, anyone within the catchments was deemed to have the same access, but this seems less problematic at very small scales. This paper showed that it is possible to conduct small scale analysis and this could be of particular importance to Tenovus when assessing their contributions to different cities or smaller areas.

Wang and Luo, (2005) examined primary healthcare accessibility in Illinois, USA using a combination of spatial and aspatial data within their FCA models. By conducting a 2SFCA analysis followed by a factor analysis to group sociodemographic variables the study aimed to provide a new way for key stakeholders to identify health professional shortage areas. The study had the same limitations as many of the 2SFCA papers in that it assumed everyone within a catchment had the same level of access, but by combining the spatial and sociodemographic results it provided a real insight into problem areas. The authors believed this type of analysis could be of benefit for healthcare provision planning and that the method is scalable to larger or smaller areas. Their study highlighted the ability of 2SFCA methods, used alongside other analytical approaches, to provide deeper insights into healthcare provision.

Luo, (2004) presented a GIS method for identifying primary healthcare physician shortages in Northern Illinois, USA. The aims of the paper were to showcase the 2SFCA method to key stakeholders, and to show that the greatest variability occurred at smaller scales, as well as recommending that finer resolution data should be used in healthcare allocation studies. The results showed that the greatest variability between population to physician ratio was at a resolution of less than 20 miles. The authors stated similar issues with the method; there was an edge effect, and the method did not account for any variance in accessibility within catchments.

This paper added further weight to the method when compared to the more commonly used accessibility methods by healthcare providers, and is of benefit to this study by highlighting the importance of spatial data resolution within such models.

Wang, (2011) studied accessibility to primary healthcare for different types of immigrants in Toronto using spatial and aspatial data, by focusing on the access immigrant populations had to physicians who could speak their language. The 2SFCA study showed that there was a cluster of supply in the centre of the city. When the immigrants were assessed, there were mixed results with some having better access than others, and there was no consistent pattern between the length of time the immigrant population has been established and accessibility. The authors noted some limitations to the study, such as equal access within a catchment, but they also suggested the fact that some assumptions were made in the study which may not be correct. The study assumed that the population would prefer to speak with a doctor in their mother tongue, and this may not have been the case. The authors proposed some points for future research; getting actual data from physicians describing their patients as well as obtaining a better understanding of the travel habits of patients. This in turn highlighted the need for the types of multi-modal approach included in subsequent implementations of the FCA approach (Langford et al., 2016).

This is a small sub-set of papers which used 2SFCA to investigate access to healthcare which showed the potential benefits of such an approach over ‘traditional’ measures of accessibility. The papers highlighted applications at a variety of spatial scales, and demonstrated how versatile the method is and how it can be utilised alongside other data to provide a rich understanding of healthcare accessibility. These papers highlighted that although this method has flaws it is possible to use it to plan healthcare provision, and provides important context for the approach taken in the present study.

2.5 Enhanced Two Step Floating Catchment Area (2SFCA) techniques

Luo and Qi, (2009) addressed some of the limitations of the standard 2SFCA method and presented the enhanced 2SFCA (E2SFCA) method. Their approach introduced distance decay which addressed the fact that, for example, it is less desirable to drive for 25 minutes to a healthcare provider than it is to drive for 5 minutes. To do this the catchment was split into three different time zones each having its own weight, with distance decay being accounted for in each calculation. The results were more intuitive as they provided more variance within catchment areas.

Luo and Qi (2009) described the method as follows:

Step 1: Calculate provider to population ratio Rj at each location j. For each j compute 3 travel time zones, search k within each zone:

$$R_j = \frac{S_j}{\sum_{k \in \{d_{kj} \in D_r\}} P_k W_r} = \frac{S_j}{\sum_{k \in \{d_{kj} \in D_1\}} P_k W_1 + \sum_{k \in \{d_{kj} \in D_2\}} P_k W_2 + \sum_{k \in \{d_{kj} \in D_3\}} P_k W_3} \quad (1)$$

Step 2: Calculate the spatial accessibility index Afi for each population site i:

$$A_i^F = \sum_{j \in \{d_{ij} \in D_r\}} R_j W_r = \sum_{j \in \{d_{ij} \in D_1\}} R_j W_1 + \sum_{j \in \{d_{ij} \in D_2\}} R_j W_2 + \sum_{j \in \{d_{ij} \in D_3\}} R_j W_3 \quad (2)$$

Notation key

Sj: medical capacity at each provider j

Rj: provider-to-population ratio

Pk: population site k

Dkj: travel time between k and j

Afi: Spatial accessibility index of each population site j

Dij: travel time between i and j

Wr: Gaussian weight for each zone

Luo and Qi (2009) conducted a case study that compared the accessibility outputs from both an E2SFCA and a standard 2SFCA approach in the northern Illinois region, in order to evaluate primary care physician accessibility. Their findings suggested that the implementation of E2SFCA models resulted in more detailed patterns of accessibility because E2SFCA differentiated between accessibility within catchment areas and was able to provide more insight than the standard 2SFCA model. Although this method addressed one of the methodological problems associated with 2SFCA, it was not able to address all of them.

E2SFCA has been used in multiple studies investigating different subject areas in different parts of the world, the studies highlight the versatility of the method and offer guidance on when best to use it. Dewulf et al., (2013), used different methods to investigate accessibility to primary healthcare in Belgium. The paper focused on comparing the most commonly utilised physician-to-population ratio with some more advanced methods. The study was a nationwide study of Belgium using supply-side data for over 10,000 active physicians (those who have over 500 patient contacts per year) and the 10.8 million inhabitants of Belgium as a demand-side variable. Population-to-physician ratios, mean distance to nearest three physicians and distance to nearest physician along with cumulative opportunity, 2SFCA and E2SFCA, were used as access measures. The study showed that the E2SFCA method was preferable to 2SFCA as it utilised the distance decay function, but that it mostly highlighted suburban and rural areas. This study highlighted the use of FCA at a national level and showed its strengths when compared to less advanced methods. The authors highlighted that there were some limitations to the study namely that the use of population centroids may be problematic in some research contexts, people may have visited a physician from a work address, and home addresses were all that were considered; and that the catchment size should perhaps have varied depending on whether it was in a rural or urban setting. This paper did highlight the efficacy of this method at a national level, and highlighted the benefit of such 'enhanced' models.

Pan et al., (2015) suggested using E2SFCA to measure access to hospitals in Sichuan Province, China. The commonly used method to measure accessibility in the region is the population to provider ratio. As Dewulf et al., (2013) showed in their study, the population to provider ratio is not the most advanced technique, and this study found that there were large parts of the population with limited access. Of the 80 million people in the province, 5.5 million lived more than 2 hours from the nearest hospital which put a large number of people at high risk. The authors suggested the need for more follow-up research in this area and called for further studies to guide stakeholders to use more advanced accessibility techniques to identify gaps in provision.

Moving the method forward, Kanuganti, Sarkar and Singh, (2016) used the E2SFCA method to evaluate healthcare accessibility in 5 districts of Rajasthan, India. Although they used the

E2SFCA method they applied different impedance functions for different population types such as sigmoidal functions for areas of high population and decline functions in lower population areas. The case study used the method of least squares regression to fit different functions to the different regions, and this provided the basis for the different function recommendations. Their study drew attention to the potential importance of the form of the distance decay parameters used within such FCA models on the resulting accessibility surfaces.

Using the method to investigate inequity within accessibility Wang and Pan, (2016) investigated the inequity of hospital accessibility between ethnic minorities and the overall population in Sichuan Province, China. The study used the E2SFCA analysis to show that the current population to provider ratio was not as good at describing access to healthcare. The paper identified 3 problems facing ethnic minorities; healthcare supply was clustered into small areas; the facilities were mainly government funded and more primary and tertiary hospitals were needed. By using E2SFCA it was possible to highlight the inequity within the region and make recommendations to stakeholders. This study highlighted the advantages of using a more advanced method and showed some interesting findings alongside delivering a more complete understanding of healthcare accessibility.

Cao, Stewart and Kalil, (2016) provided an investigation into end stage renal disease in Midwestern USA. The study analysed spatial and aspatial data to see if variations in end stage renal disease occurred between White, Black and Native American and rural/urban populations. The use of E2SFCA helped to identify that although accessibility to treatment was similar between white and black groups, black transplant rates could be as much as 73% lower. The study highlighted the need to understand the travel time of the demand and what different distance decay functions should be used depending on the area (rural or urban); and showed that accessibility to transplant centres had no significant bearing on end stage renal disease hot spots. In this case another variable, kidneys which are available for transplant, had a significant impact on the results.

Similarly, Wang and Pan (2016), Vadrevu and Kanjilal, (2016) assessed accessibility to healthcare for pregnant women in the Sundarbans region of India using the E2SFCA method. This study introduced an area-based socioeconomic score to measure inequity between the population. By combining the accessibility and socioeconomic scores the study showed that many places were underserviced, and that there was not an equitable provision of maternal healthcare. The study suggested that there were 3 keyways to improve healthcare provision; equitable distribution, efficient transport and a focus on basic infrastructure in the region. This paper showed that it was possible to use E2SFCA models with other measures to provide a richer understanding of accessibility.

These papers showed that E2SFCA can be used to study accessibility and a wide range of services and it provides a more robust alternative to the commonly used population to provider ratio. The studies operated at both regional and national levels and were used to investigate not just healthcare accessibility but to evaluate the accessibility of different parts of the population. Different impedance functions were used in conjunction with this method to provide more accurate results in different circumstances and most of the papers are championing the use of more advanced accessibility methods to improve the wellbeing of different parts of the population. E2SFCA addresses some of the issues highlighted in the 2SFCA method by adding distance decay within the catchments and by allowing different functions to be used in different situations.

2.6 Major Advancements in Two-step Floating Catchment Area (2SFCA) models

Since its initial formulation, a number of enhancements to the basic 2SFCA approach have been proposed (summarised in Table 2.1). These addressed perceived weaknesses or issues highlighted with the 2SFCA method, and tried to optimise the method for different scenarios and data. These advancements allowed the method to provide more realistic results in differing situations across different geography and populations. These are some of the most widely adopted advancements, and further examples are provided in Table 2.2.

Method	**Summary**
Kernel Density 2SFCA (KD2SFCA)	This method removed the steps that were prevalent in other methods by using a continuous distance decay function across the entire catchment. It also used a principal components analysis on various socioeconomic variables to provide an index of healthcare needs. See (Dai and Wang, 2011; Polzin, Borges and Coelho, 2014).
Optimised 2SFCA	This method weighted the calculation of travel times. Using a mix of weighted supply and demand data the authors calculated demand

	based on the proportion of medical users, and aggregated the demand counts at primary healthcare sites. The authors stated that this method is only valid at small distances and does not scale. See Ngui and Apparicio, (2011).
3SFCA	This method used a selection weight to address the overestimation of demand. It added a selection weight to the surrounding supply sites in both catchment calculations. See Wan, Zou and Sternberg, (2012).
Variable Catchment Size 2SFCA (V2SFCA)	This method used small increments in catchment size until a predetermined supply to demand ratio was met. It attempted to address the arbitrary nature of catchment size selection and make the method more effective over larger and more rural areas. See Luo and Whippo, (2012).
Huff Model Based 2SFCA	Integrated the Huff Model with FCA to address how demand is calculated. It used a probability value to predict higher demand locations and was adjusted for population demand. See (Luo, 2014).
Multi-Modal 2SFCA (MM2SFCA)	This method used more than one transport network to better understand how the supply side accesses the demand. This could be including a networked distance for driving and for public transport. See (Langford, Higgs and Fry, 2016).

Table 2.1 Table of advancements in 2SFCA methodology

2.6.1 Kernel Density 2SFCA (KD2SFCA)

Dai and Wang (2011) introduced the KD2SFCA method while investigating accessibility to healthy food opportunities in Mississippi, USA. The kernel density function was used in a similar way to the original method, but the kernel bandwidth was set to the same size as the catchment and the score was discounted as it got closer to the edge of the catchment, and became zero after that distance was met. The study showed that KD2SFCA better described accessibility in rural areas compared to the 2SFCA method which suggested good accessibility in rural areas; applying the KD2SFCA method in contrast suggested poor accessibility.

The study authors claimed three contributions of their research which involved:

- Using factor analysis to consolidate aspatial factors and using these to investigate different groups within the population.
- Proposing an advanced measure to capture spatial accessibility.
- Separating the spatial and aspatial factors in accessibility.

Polzin, Borges and Coelho, (2014) aimed to improve the standard KD2SFCA method and investigate access to healthcare facilities in Portugal. The authors attempted to optimise the KD2SFCA method to analyse the access to healthcare facilities and highlight less empowered parts of the population. The enhancements highlighted more areas than the original KD2SFCA method by including additional access dimensions.

Advancing the method Tan et al., (2018) investigated the rural-urban accessibility of hospitals in Australia using a modified KD2SFCA (MKD2SFCA). The study calculated a more realistic demand point using satellite data as in suburban and rural areas, the population was not described well using other methods. The study also used a Gaussian kernel as the density function as it had a gradual rate of reduction. The study found that although those in remote areas had similar accessibility rates, factors could have large implications. The access to emergency care was much diminished in rural regions, and the weather could have a significant impact on certain areas.

These studies gave a more detailed insight into accessibility but this came at a cost, the methods were much more intensive and required more resources to compute. Separating spatial and temporal data offered the ability to gain considerable insight but introduced more complexity into the research design which may not have been appropriate for some types of users or application areas.

2.6.2 Optimised 2SFCA

Ngui and Apparicio (2011b) attempted to improve the 2SFCA assumption that all services within the same catchment were equally accessible, and that the population within these catchments used the services equally. To do this the authors weighted the supply by the number of physicians working in each location, and the demand by the percentage of people who use each facility during a 12-month period. A study was conducted in Montreal, Canada which used ArcGIS to compare the 2SFCA method and the Optimised 2SFCA method. The results showed that the improvements gave a more accurate overview of accessibility at small scales (500m or less) and that there was no significant difference as the scales increased.

This method offered great insight into what can be done with the 2SFCA foundation, the method is only really successful at very small scales and requires a great amount of data to be properly implemented. This may not be feasible to use at large scales and as such may be of limited use in this project.

2.6.3 3SFCA

Wan, Zou and Sternberg, (2012) proposed a novel solution to the overestimation of healthcare demand. This method built upon the E2SFCA method and addressed the two limitations with gravity and E2SFCA models which were; over estimation of demand for some healthcare facilities and secondly, the impedance coefficient that was used. To address the issues, they assigned a travel-time-based competition weight for each supply-demand pairing and used this in the demand calculation. A case study was conducted using access to primary care physicians in the Texas area of the USA. The case study showed that it was effective in reducing the overestimation of demand but although there were advantages the authors acknowledged that there were a few points which needed refinement, such as the way it divided the catchments into different zones. They acknowledged this could be done in a different way to better represent the population and the catchment size, which could be changed to give a more accurate representation.

Advancing the method, Ranga and Panda, (2014) completed a study on access to in-patient healthcare in rural India. The authors used questionnaire data from supply and demand sides as well as hand held GPS data to establish the positions of 34 in patient healthcare providers and 3,685 households, 3 sites were used for this study. A Gaussian distance decay was applied and 3 bands were used to split the catchment areas at 2, 5 and 10km. As many of the data points did not connect to a road network Euclidean distance was used for the analysis, after comparisons between networked and straight-line distances showed a significant relation. The authors welcomed the new advancements within the FCA methodology but suggested that most are tested in the developed world, and that they are often under evaluated in the third world where road networks and infrastructure vary dramatically. The authors found that healthcare facilities were based in urban or semi-urban areas to maximise profit and this pushed poorer people to worse conditions than they were already experiencing. The paper suggested that similar limitations were prevalent with this method, namely in relation to the aggregation level of the spatial data and the boundary definitions that could obscure patient flow.

2.6.4 Variable Catchment Size 2SFCA (V2SFCA)

V2SFCA (Whippo,2012) intended to address the limitations in previous 2SFCA models around how the catchment size was calculated, by utilising a base population threshold and physician-to-population ratio to establish the size of each catchment. This was done incrementally in the first step of the calculation, and the catchment size was adjusted until the criteria was met. Each catchment was split into 3 zones and a Gaussian function was used to calculate the distance decay. A case study was completed in Illinois, USA to measure the effects of the new method and the results showed more detail in urban areas and “intuitively reasonable” patterns in rural areas.

Ni et al., (2015) suggested an enhanced V2SFCA (EV2SFCA) to address some of the limitations of the V2SFCA model. They suggested that V2SFCA probably “calculates inaccurate supply catchments” which affected the results. To address this, they recommend a method which accurately calculated catchment size and minimised the errors from mismatched supply and demand catchments. This was done by dynamically summing the supply to demand until a supply to demand threshold was met. A case study in China was used and it highlighted some areas with very low accessibility which needed to be addressed. This method offered a great advancement to the overall methodology where in most cases an arbitrary number was assigned to the catchment size, this could lead to drastically different results depending on how it was set. Having a way to accurately calculate this would help to improve the quality of the method and provide more insightful and realistic results. This project had access to actual demand side travel times and distances so they would be used as a base for the catchment making this method unnecessary.

2.6.5 Huff Model Based 2SFCA

Luo, (2014) proposed the Huff Model Based 2SFCA to moderate the over or under estimation of demand (population) on healthcare services. The method also included a continuous distance decay function which removed many of the arbitrary bands that the other methods used within catchments. To test the method a case study was completed in Missouri, USA investigating access to primary care physicians using ArcGIS. The study showed that the equal access assumption within 2SFCA leads to overestimation of accessibility, and that this method showed more variations within catchments.

2.6.6 Multi-Modal 2SFCA (MM2SFCA)

A multi-modal 2SFCA was developed by Langford et al (2016), and applied to a study concerned with measuring access to primary healthcare in South Wales using different types of transportation network. The methodology aimed to understand the differences in accessibility between different segments of the population by assessing how those who used different types of transport had varying levels of access to services. Using separate road networks for public and private transport it was possible to compare the accessibility results of different cohorts within a population. The case study showed that the access of bus users was overestimated compared to car users, and that car user accessibility was potentially underestimated due to the lower access levels of the bus users. The authors noted that the changes in results varied dramatically by adjusting the catchment area used, and that this needed to monitored by those implementing this solution. Their findings also drew attention to the limitations of assuming all individuals were able to access services via private (i.e., car) means of transport.

More recently, Tao et al., (2018) suggested improvements to the MM2SFCA model by incorporating traffic API's to increase travel time accuracy. A case study was conducted in Shenzhen, China to gain further knowledge of the effects of incorporating multiple modes of transport. The study included data from a Badu Map API as opposed to a standard road network with the arbitrary speeds assigned to each part of the network. The authors suggested that the disparity of results between the private transport sub-group and the public transport sub-group was large, and was not visible in traditional 2SFCA approaches which assume single-mode forms of transport. The paper suggested that one way to drastically reduce inequality of access to healthcare would be to improve the public transport networks. The authors noted that the car ownership data they had was at a city scale and that it could be amended to include sub-district level data to get a deeper understanding of the disparities at a finer scale.

The ability to assess disparities in transport is one of great importance and provides great insights that are difficult to observe without different transport models being accounted for. Research officers within Tenovus confirmed that the use of public transport for chemotherapy is very uncommon, and public transport to and from appointments would be unadvisable so whilst it may not be necessary to incorporate multiple modes of transport within the overall approach taken in this study it could be implemented in further iterations of the software solution (see Chapter Eight).

2.6.7 Timeline of FCA methodological advancements

Figure 2-1 shows a timeline of advancements which are described in more detail in tables 2.1 and 2.2, this represents a summary of what was discussed in the previous sections. The speed in which methods have been advanced to enhance the initial formulation of the 2SFCA approach shows the perceived advantages of the application of such tools and provides further weight for investigating the potential use of FCA models in the present study.

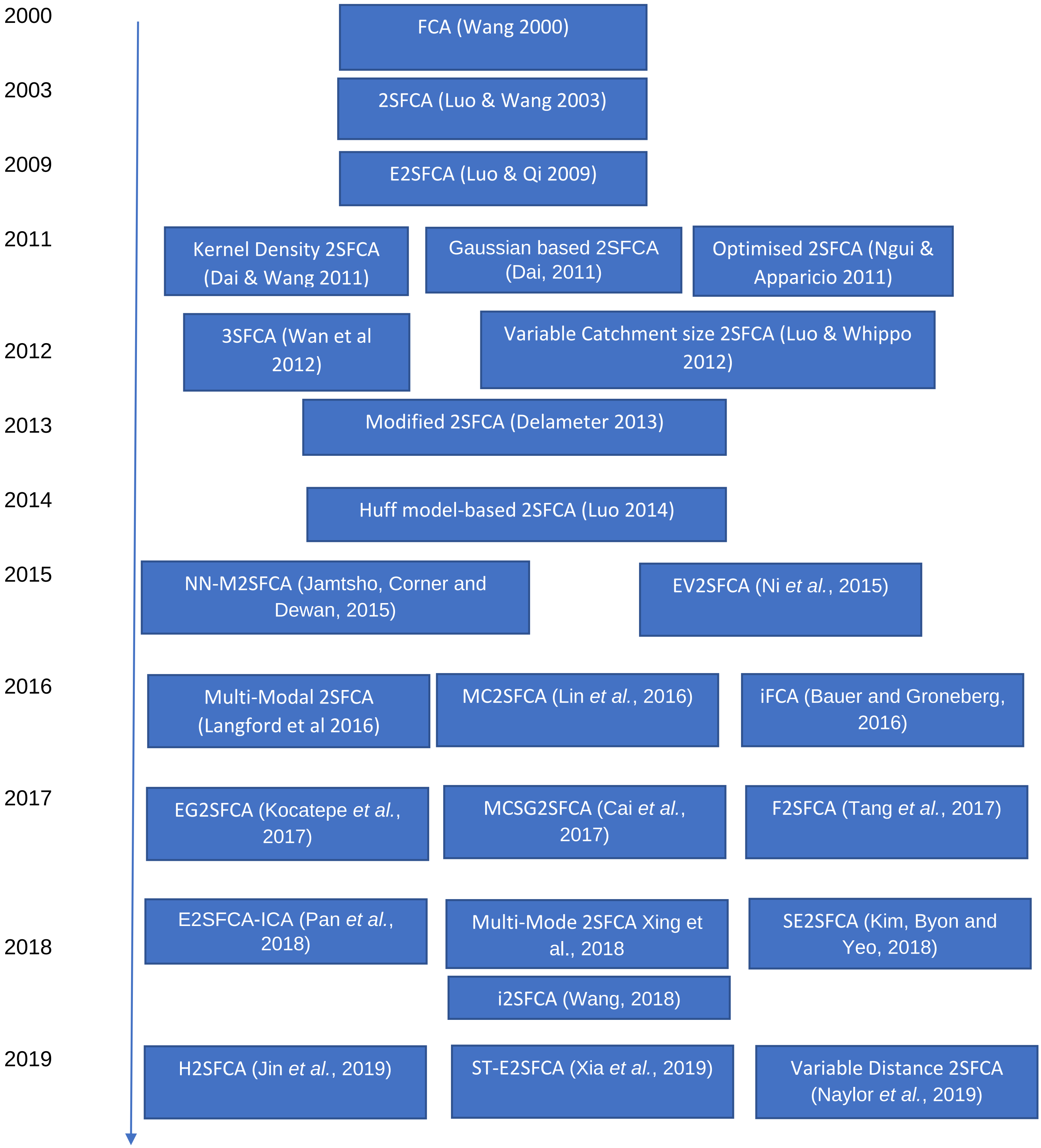


Figure 2-1 Timeline of key 2SFCA advancements

Alongside these main enhancements, most papers enhanced the basic 2SFCA model to suit the types of data being used and the particular study context (Tables 2.1 and 2.2). A large number of papers introduced aspatial supply or demand-side data, to get a better understanding of potential inequalities in accessibility for particular segments of the population. This is particularly interesting in smaller catchment sizes and when ethnicity or socio-economic status is being investigated (Cabrera Barona and Blaschke, 2015; Jones et al., 2016; Lin et al., 2016). Using additional aspatial data was a good way of investigating potential inequalities for sub-groups within the overall population. Figure 2.1 shows a timeline of advancements

Table 2.2 highlights most of the methodological advancements within the 2SFCA method. The table briefly describes the timeline and the aims of the proposed enhancements to the basic 2SFCA methodology that have been investigated in the last two decades.

Method	**Author**	**Aim**
Gaussian based 2SFCA	(Dai, 2011)	Used a Gaussian based distance decay to more appropriately assign distance decay weights. Used a distance decay with no bands. Needed accurate data to set the Gaussian curve.
Hierarchical 2SFCA (H2SFCA)	(Jin et al., 2019)	This method introduced different distance attenuation functions to represent 3 levels of healthcare service provision across multiple facilities based upon the different types of service each hospital uses. The authors noted that the population distribution should be investigated as it has an effect on service distribution.
NN-M2SFCA	(Jamtsho, Corner and Dewan, 2015)	This method aimed to provide accessibility scores for entire countries by using spatial and temporal information. Demand sites were identified using cluster analysis. This method had a higher complexity of calculation compared to the original method and required historical data. Difficult to compare with other methods due to the lack of catchment sizes.
Spatio-temporal	(Xia et al., 2019)	To include temporal data alongside the spatial data by using large scale GPS data. This method provided a

E2SFCA (ST-E2SFCA)		detailed illustration of the spatio-temporal accessibility. Can be difficult to gather the required data and increased the complexity of the method.
Variable - Distance E2SFCA	(Naylor et al., 2019)	This method attempted to investigate accessibility at a national scale and did so by introducing several measures to the E2SFCA method. By using the Variable-Distance E2SFCA method at zip code level and then analysing the results at a small level the results were analysed at a national level using generalised linear models. This method was more difficult to calculate than other methods but had benefits for large scale studies.
Multi-criterion 2SFCA (MC2SFCA)	(Lin et al., 2016)	This method can assist in resource allocation while considering multiple risks such as age of population, population density and land use classification, by using multiple methods this method attempted to highlight disparity in urban settings and deficiencies in rural settings.
E2SFCA – Integrated catchment area (E2SFCA-ICA)	(Pan et al., 2018)	E2SFCA-ICA uses actual data from GPS and includes it in the E2SFCA method. By using integrated access probability; an integrated catchment area was created which reduced the differences between model and data-based catchment areas and highlighted the core catchments.
iFCA	(Bauer and Groneberg, 2016)	iFCA introduced a different way of computing the FCA methods with a variable distance decay function which used the median and standard deviation to compute catchment sizes, based on the individual population and supply of an area, removing the need to assign a catchment size.
Multi-Mode 2SFCA	(Xing, Liu and Liu, 2018)	Multi-Mode 2SFCA used public green spaces to introduce a new method classifying the spaces based on the services provided and the demands of the population. This method showed slightly more realistic results than the traditional 2SFCA method.

SE2SFCA	(Kim, Byon and Yeo, 2018)	Specially designed 2SFCA for the City of Seoul in South Korea, as it is different from most others in the world with higher population density and smaller travel distances to healthcare providers.
i2SFCA	(Wang, 2018)	The inverted 2SFCA method proposed to assess crowdedness of facilities. This was done by using the supply capacity as a weight and reversing the 2SFCA process by initially measuring competition for supply and then in step 2 summing up the demand.
EG-2SFCA	(Kocatepe et al., 2017)	Empirical-Gaussian 2SFCA proposed to analyse the exposure to severe injury crash hotspots on a road network with a socioeconomic weighting. This measure highlighted certain age ranges being more exposed to these sites than others.
MCSG2SFCA	(Cai et al., 2017)	Multi catchment size Gaussian 2SFCA investigated different access to private and public aged care facilities. To do this it divided the facilities into different categories based on capacity and a catchment time assigned based on different gathered statistics.
F2SFCA	(Tang et al., 2017)	This method introduced aspatial dimensions into the 2SFCA calculation to mitigate some of the problems with the 2SFCA method by estimating the flow of people using actual data. The main advantage of this study was at small scales where it provided more realistic results.
EV2SFCA	(Ni et al., 2015)	Aimed to address the over estimation of the supply calculation in the V2SFCA. More accurately represented the supply and created a more calculated catchment area.

Table 2.2 Aims of the proposed methods

2.7 Applications of 2SFCA

This section shows the different ways in which 2SFCA has been used and the types of studies that have adopted a floating catchment area approach. Although the majority of studies initially focused on their application in health geography, such methods are increasingly being used to investigate accessibility to a wider range of facilities/services. In particular, they have been used to investigate access to:

- Sport facilities (Langford et al., 2018).
- Green spaces (Gu, Tao and Dai, 2017; Wang and Wang, 2018; Wu et al., 2018; Xie et al., 2018; Xing, Liu and Liu, 2018; Li et al., 2019).
- Healthy food opportunities (Dai and Wang, 2011; Chen, 2019; Wiki, Kingham and Campbell, 2019).
- Schools (Ye et al., 2018).
- Libraries (Guo, Chan and Yip, 2017).
- Pharmacies (Wang and Ramroop, 2018).
- Employment (Dai et al., 2018).
- Day care centres (Fransen et al., 2015).
- Transport systems (Langford et al., 2012; Naranjo-Gómez et al., 2019).

Within health geography, one of the most studied areas is access to primary healthcare physicians (Luo & Wang, 2003, Wang & Luo, 2004, Luo, 2004, Wang et al, 2008, McGrail & Humphreys, 2009, Luo & Qi, 2009). More recently such models have been used as part of wider studies concerned with investigating the role of access on patterns of cancer diagnosis, treatment and survival and in examining potential associations with other healthcare outcomes. It has also been used to study variations in access to other types of healthcare facilities such as:

- Dentists (Jones et al., 2016; Nasseh, Eisenberg and Vujicic, 2017; Shin and Ahn, 2018).
- Physiotherapists (Shah, Bath and Milosavljevic, 2015).
- Mental healthcare services (Ngamini Ngui and Vanasse, 2012).
- Maternity services (Song et al., 2013; Vadrevu and Kanjilal, 2016; Yin, 2018).

Utilising different 2SFCA methods it was possible to study accessibility at different scales. For example, its application to extensive geographical areas has been investigated (McGrail and Humphreys, 2009a), whilst its use in very local scale studies has been explored Ngui and Apparicio, 2011b). Studies have been conducted in a variety of different countries demonstrating its generality to alternative geographical environments. For example;

- North America (Ngamini Ngui and Vanasse, 2012; Wan et al., 2012, 2013; Mao and Nekorchuk, 2013; Luo, 2016; Khakh, Fast and Shahid, 2019; Naylor et al., 2019; Zahnd et al., 2019),
- Europe (Polzin, Borges and Coelho, 2014; Bauer and Groneberg, 2016b; Gao et al., 2016; Langford, Higgs and Fry, 2016; Siegel et al., 2016; Bauer et al., 2017, 2018; Naranjo-Gómez et al., 2019).
- Asia (Choosumrong, Raghavan and Bozon, 2012a; Yu, Tan and Ruan, 2012; Pan et al., 2015; Vadrevu and Kanjilal, 2016; Ma et al., 2018; Tao et al., 2018; Xing, Liu and Liu, 2018; Chen and Yeh, 2019; Xia et al., 2019; Ranga and Panda, no date).
- South America (Polo, Acosta and Dias, 2013; Rocha et al., 2017; Cristina Rocha-Brischiliari et al., 2018).
- Australia and New Zealand (Scott et al., 2006; McGrail and Humphreys, 2009a, 2009b, 2014; KC, Corcoran and Chhetri, 2018; Wiki, Kingham and Campbell, 2019).
- Africa (Tansley et al., 2016).

This list is not exhaustive and there is a burgeoning literature on the application of such tools; The flexibilities of the method make it ideal for a project such as this concerned with examining spatial patterns of accessibility to different types of cancer service provision.

2.8 Applications of Floating Catchment Area Techniques in the management, treatment and planning of cancer services

In addition to general healthcare accessibility, 2SFCA is utilised extensively within cancer accessibility research. This section highlights papers that have specifically investigated the effects of accessibility in relation to cancer services and provides a synopsis of key papers as well as outlining those methods that have been primarily used to date. Most of the papers are looking at primary care access and links to late-stage diagnosis.

Wang et al., (2008) investigated the link between late-stage breast cancer diagnosis and access to primary healthcare and mammography screening services in Illinois, USA. Using 2SFCA the study showed that poor accessibility to primary healthcare increased the late-stage breast cancer diagnosis, particularly noting that those living amongst a high concentration of disadvantaged people were at the most significant risk. There was no significant link between mammography screening accessibility and late-stage breast cancer diagnosis in this study. The authors pointed out the limitations with the data they used but, this study showed important results for healthcare policy makers.

Rocha-Brischiliari et al., (2018) conducted a study in Parana, Brazil which assessed the access and socioeconomic disparities across 399 cities. A 2SFCA method was used in conjunction with other techniques to measure accessibility to cancer services across the study area. The study found that breast cancer mortality rates were positively associated with access to cancer services and negatively associated with illiteracy. This study showed how useful the 2SFCA method can be when used in conjunction with other data to obtain a more detailed understanding of a specific issue. Zahnd et al., (2019) completed a study investigating access to mammography in 8 states of the USA. The study used E2SFCA to compare impoverished regions with high levels of black residents with low screening rates to other areas in the region. The study showed that even though there were low utilisation rates of the services, the spatial accessibility did not vary across regions. Xu et al., (2017) used the 2SFCA method to evaluate access to NCI Cancer Centres in the USA. The study investigated the disparities across demographic and urban areas. It showed that access to the NCI Cancer Centres was greater in large metropolitan areas and suggested that populations under the poverty line were disproportionately concentrated in lower accessibility areas, and that non-Hispanic white had the lowest geographical accessibility to these services. Kanuganti et al., (2016b) used 2SFCA to conduct a case study in the Alwar district of Rajasthan. The district had 2,217 villages and an 8km catchment area was utilised. The accessibility map highlighted areas with poor accessibility, and the authors noted that this could be used by policy makers to improve accessibility within the region.

A new framework was suggested by Gu, Wang and McGregor, (2010) to measure accessibility (modified 2SFCA); building on the 2SFCA method the authors added a distance factor and a Huff-based competitive model. They conducted a case study on breast cancer screening services in Alberta, Canada and the results showed that it was beneficial when planning where to locate preventative healthcare facilities. Donohoe et al., (2016) evaluated access to mammography within Appalachia, USA, and compared multiple 2SFCA variations. They highlighted several regions within Appalachia which had poorer accessibility and recommended using the spatial access ratio to minimise the differences between the different 2SFCA variations. McLafferty et al., (2011) conducted a study into rural-urban inequalities in late-stage breast cancer diagnosis for both the overall population and the African-American population, in two time periods 1988-92 and 1998-2002. They did this by using several techniques including 2SFCA for the spatial aspect of the paper. The results showed that there was no significant rural-urban inequality for the overall population; there was a decline in late-stage breast cancer diagnosis between the two time periods with bigger decreases in higher population areas. There was a large disparity between the overall population and the African-American population in both time periods.

A study was conducted using E2SFCA and aspatial methods to investigate the links between accessibility and colorectal cancer in the Texas, USA region. The study used three weights for the different oncologists in each catchment area, and showed that accessibility to oncologists made a difference in rural areas, but not in urban areas. The aspatial data showed that disadvantaged groups had more chance of dying from colorectal cancer (Wan et al, 2012). An evaluation of current methodology from the gravity model through 2SFCA, including KD2SFCA which discussed different distance decay functions. The paper discussed the need for greater utilisation of spatial and aspatial data, and discussed the implications it had on mortality rates using cancer in the USA as an example. The paper suggested a framework should be implemented and recommended using optimization techniques (Wang et al., 2012). Wan et al (2012) Introduce the idea of using a spatial access ratio derived from the E2SFCA spatial access index (SPAI). A case study into the accessibility of colorectal cancer facilities in Texas, USA was used to trial the proposed methods. The study used a sensitivity analysis to show that the SPAI value fluctuated with changes in the impedance coefficient, whereas the SPAR stayed stable throughout. The authors noted that if the impedance was from some basis and not an arbitrary input then the SPAI value would be valuable and could show more detailed results using E2SFCA and a factor analysis to assess the effect of spatial and aspatial accessibility to oncologists in Texas, US. The study examined the effect of accessibility with regard to colorectal cancer survival, the E2SFCA method used a Gaussian function and a three-step catchment area. Cox proportional hazard regression was used to analyse the association between healthcare accessibility and colorectal cancer survival. The factor analysis and principal components analysis were completed, and nine aspatial data sets were included in the study; including household income, home valuation and high school education. The results showed that several minority groups were more likely to die from colorectal cancer, and that spatial access to oncologists had a significant effect on colorectal survival in non-urban areas, but not in urban areas themselves (Wan et al, 2012).

Shi et al., (2012) used 2SFCA to evaluate accessibility to cancer care facilities in the USA. The study suggested that the travel impedance threshold was the crucial factor when conducting a 2SFCA calculation as this was a special case of the gravity model. It also suggested that different choices in distance decay should not affect the overall pattern when used. This was a very large area for a 2SFCA study and it showed that there were two main regions with lower accessibility, the east and west of the USA. The west had deeper pockets of poor accessibility, whereas the east had larger areas with slightly lower accessibility ratings. Donohoe et al., (2015) compared the 2SFCA method to traditional accessibility methods in the Appalachia region of USA to assess the impact of access to mammography centres and primary care centres on late-stage breast cancer diagnosis. The study showed that accessibility was not suitable and that the 2SFCA method offered the “greatest predictive validity”.

Donohoe et al., (2016) evaluated the access to mammography within Appalachia, USA and compared multiple 2SFCA variations. They highlighted several regions within Appalachia which had poorer accessibility and recommended using the spatial access ratio to minimise the differences between the different 2SFCA variations.

These are just a selection of the studies that have used floating catchment area approaches to investigate different aspects of cancer care accessibility, which encapsulates everything from regional patterns of healthcare accessibility to associations with stage of diagnosis at smaller spatial scales. Such models are also being used to help select where new resources should be placed and how to optimise current cancer services which provide an important context for the present study concerned with developing open-source solutions to examine spatial patterns of accessibility to cancer services in Wales.

2.9 Interpreting 2SFCA results

The studies mentioned in chapter 2 and the results section of this thesis rely heavily on visualisation for interpretation of the results of 2SFCA studies (Shi et al., (2012) and McGrail & Humphreys (2014)). The visualisations provide an overview where it is possible to observe a large amount of data in an easy to understand format. McGrail & Humphreys (2014) state that 2SFCA results can be interpreted like provider to population ratios to 100,000 so a score of 0.000200 would be 200 providers per 100,000. Many studies use other methods of evaluation on the data and attempt to look for patterns and observations but the higher the score an area receives the higher the level of accessibility within that area compared to those accounted for in the study. The numbers will change between datasets and it is the distance between the scores that is of interest and can assist in making comments on accessibility. Papers have used alternative methods to compare the FCA methods such Donohoe et al., (2016) used Spearman Correlations to compare different FCA methods with non-FCA methods, Dewulf et al., (2013) used a two tailed Pearson correlation test, Pan et al., (2015) calculated the correlation coefficient and paired t-tests and Luo & Wang (2003) use the standard deviations to analyse the results. There are multiple ways analysis can be done on FCA results, in this study the numbers are interrogated and maps are checked to ensure they appear intuitive.

2.10 The use of Open-Source GIS technologies

Studies concerned with applying GIS tools to measure access to healthcare services in general, and cancer facilities in particular, tended to be undertaken using proprietary GIS packages such as ArcGIS. However, there is an increasing movement concerned with developing open-source solutions to address these real-world issues and this is described in this section. This will briefly review the history of GIS and the major advancements that have occurred. GIS has assisted in how we study the world; and with its increasing capabilities and ever more powerful computers available, has enabled us to be able to compute larger and more complex calculations. Since the inception of GIS in the 1960s the capabilities of the different systems have increased enabling geographers, amongst many, to increase the scope of their work and improve the accuracy of the results.

GIS can be defined in various ways (Clarke, 1986; Maguire, 1991). Many of these discuss how it is a software and how it is used. For the purpose of this GIS can be defined as “a technology, as computer-based systems for integrating and analysing geographic data”. Goodchild, (1995) set out 4 functions to distinguish GIS as a technology; the ability to store, compute and display spatial relationships, the ability to analyse spatial and attribute data, the ability to integrate spatial data from different sources and store many attributes of objects. Goodchild, (2018) discussed the development of GIS through 1960s projects including the Canada Geographic Information System (CGIS), a project at Northwestern University, and work at the Experimental Cartography Unit in the UK. The computing restraints in the 1960s did not allow for the type of GIS we see today, and the features were limited. Khan and Mutahar Aaqib, (2017) showed the timeline as “custom programs created by government agencies on mainframe computers within the Nineteen Sixties, to non-public desktop based mostly software system within the Eighties, to integrate mostly net based solutions within the 2000s”.

GIS can perform a great number of tasks and some GIS have a wider range of spatial analytical functionality. One of the key developments in GIS has been the ability to use current functionality and libraries within the GIS to create software that is fit for a specific need. For example, within QGIS there is a plugin builder which allows developers to use Python scripting and create a plugin which does a desired task (QGIS 2 Cookbook, Mandel et al, 2016). ESRI provides a similar service where they can custom make a tool for a specific client's needs.

2.11 Free and Open-Source Software

Free and open-source software has been widely used in studies, but the availability of proprietary software in academic institutions has limited its widespread use. Within the healthcare accessibility studies mentioned above ArcGIS is the predominant GIS used, although there are papers where other GIS packages have been used. This section highlights the use of FOSS within academia and then focuses to show its use within healthcare accessibility studies. FOSS was born in the in the early 1980s as a response to the increasing tendency for proprietary software. Richard Stallman founded the Free Software Foundation (FSF) and in 1985 wrote the GNU manifesto. Copyright law was used to protect the developers' rights to share, and more importantly protected the developers from someone taking their source code and creating a proprietary solution (Carillo and Okoli, 2008). The FSF created the General Public Licence (GPL) in 1989. It had 10 sections stating that you can copy, modify and distribute the code, with caveats about the continued distribution of the code being subject to this licence. This has been updated with the GPL2 in 1991, and the GPL3 in 2007, which have clarified points and tried to keep the core values of the FSF.

In the late 1990s there was a new movement which saw Open-Source Software (OSS) coined as an alternative to free software. There was suspicion that big business would not adopt 'free' software (Von Hippel and Von Krogh, 2003), and in 1998 OSS was set up using similar licences but focusing on the benefits of OSS and not the morals surrounding it. Stallman says that "The terms "free software" and "open source" stand for almost the same range of programs. However, they say deeply different things about those programs, based on different values. The free software movement campaigns for freedom for the users of computing; it is a movement for freedom and justice. By contrast, the open-source idea values mainly practical advantage and does not campaign for principles. This is why we do not agree with open source, and do not use that term."

The Open-Source Initiative (The Open-Source Definition | Open-Source Initiative, 2007) defines open-source software in 10 points:

> Open source doesn't just mean access to the source code. The distribution terms of open-source software must comply with the following criteria:
>
> 1. Free Redistribution
> The licence shall not restrict any party from selling or giving away the software as a component of an aggregate software distribution containing programs from several different sources. The licence shall not require a royalty or other fee for such sale.
>
> 2. Source Code
> The program must include source code, and must allow distribution in source code as well as compiled form. Where some form of a product is not distributed with source code, there must be a well-publicised means of obtaining the source code for no more than a reasonable reproduction cost, preferably downloading via the Internet without charge. The source code must be the preferred form in which a programmer would modify the program. Deliberately obfuscated source code is not allowed. Intermediate forms such as the output of a preprocessor or translator are not allowed.
>
> 3. Derived Works
> The licence must allow modifications and derived works, and must allow them to be distributed under the same terms as the licence of the original software.
>
> 4. Integrity of The Author's Source Code
> The licence may restrict source-code from being distributed in modified form only if the licence allows the distribution of "patch files" with the source code for the purpose of modifying the program at build time. The licence must explicitly permit distribution of software built from modified source code. The licence may require derived works to carry a different name or version number from the original software.
>
> 5. No Discrimination Against Persons or Groups
> The licence must not discriminate against any person or group of persons.

6. No Discrimination Against Fields of Endeavour

The licence must not restrict anyone from making use of the program in a specific field of endeavour. For example, it may not restrict the program from being used in a business, or from being used for genetic research.

7. Distribution of Licence

The rights attached to the program must apply to all to whom the program is redistributed without the need for execution of an additional licence by those parties.

8. Licence Must Not Be Specific to a Product

The rights attached to the program must not depend on the program's being part of a particular software distribution. If the program is extracted from that distribution and used or distributed within the terms of the program's licence, all parties to whom the program is redistributed should have the same rights as those that are granted in conjunction with the original software distribution.

9. Licence Must Not Restrict Other Software

The licence must not place restrictions on other software that is distributed along with the licenced software. For example, the licence must not insist that all other programs distributed on the same medium must be open-source software.

10. Licence Must Be Technology-Neutral

No provision of the licence may be predicated on any individual technology or style of interface."

In 1991 Linux was released using a GPL, as have several major projects since; Mozilla Firefox an internet browser, PostgreSQL a database management system, WordPress a blogging platform, VLC media player, 7-zip a file archiver and Ubuntu an operating system for Linux based systems are just a few of the OSS projects.

Khan and Mutahar Aaqib, (2017) compared open-source software with proprietary software and highlighted that the documentation between Esri products and QGIS was different. There was a lack of user support offered from FOSS solutions without paying an external consultant. Training costs were also likely to be incurred as was the knowledge to install the system correctly. It was stated that there was no fee for the software, it was scalable and there were no usage limits. There were definite positive and negative points to using FOSS but it could be a very suitable solution for many organisations.

Fitzgerald, (2006) highlighted the advancements in FOSS particularly noting the differences in how the development cycle had changed from what he coined OSS 2.0. The paper suggested that FOSS is now a viable alternative with advantages over proprietary software. In this change of approach, the author noted that there were more paid developers working on projects and that organisations were paying for the development of FOSS more regularly than previously. The author highlighted the planning phase as being far more purposeful, with the planning being done by major players trying to gain an advantage as opposed to previously it being a developer with an 'itch to scratch'. This paper highlighted the change of perspective in relation to FOSS and signaled a strong future for FOSS platforms.

Within the spatial realm there are a large number of open-source projects. A large contributor is Open-Source Geospatial Foundation (OSGeo), a non-profit organisation which wants to further open technology within the geospatial sphere. Under the OSGeo banner are a variety of projects including desktop GIS systems GRASS, QGIS and gvSIG. There are several web mapping projects including MapServer, OpenLayers and GeoServer, as well as a spatial database PostGIS, which works with PostgreSQL. Many of these have been used in academic research and have helped to advance science. pgRouting is a FOSS routing library which is used to add routing functionality to the PostgreSQL/GIS database and has been used in projects in various areas of academia. It has been used too in emergency route planning (Choosumrong, Raghavan and Bozon, 2012a; Bhanumurthy et al., 2015). Studies have been completed looking to improve its performance (Sankepally and Rajan, 2018) and Yoshida, Song and Raghavan, (2010) used it in the development of a track log and points of interest system.

FOSS GIS systems are not as well utilised in papers in this thesis as the proprietary solutions, but they have been utilised in multiple studies across a range of different research areas, Steiniger and Bocher, (2009) suggested there were advantages to using a FOSS GIS over a proprietary system, and encouraged them to be more widely used. QGIS has been used to assess Open Street Map (OSM) data (Sehra, Singh and Rai, 2017) and several plugins have been created for a specific use (Jung, 2012, Nielsen et al., 2017). gvSIG is another FOSS GIS (Díaz and Anguix, 2008, Moutahir, Moutahir and Agazzi, 2012), it is widely used in Spain and there have been publications regarding additional modules (Vázquez-Rodríguez, Pérez-Risquet and Torres-Cantero, 2013), its use as an educational tool (Dorner, Scheffer and Zink, 2011). GRASS is an older GIS and has been used in studies including additional toolkits (Aw Jasiewicz and Metz, 2011) geographical modelling (Rocchini et al., 2012) and building detection from laser scanning data (Rutzinger, Höfle and Stötter, 2006). The spatial database PostGIS has been used in studies ranging from comparative studies between SQL and non-SQL databases (Piórkowski et al., 2019), extensions to the basic PostGIS functionality (Lizardo, Davis Jr and Lizardo, 2017) and data modelling (Kucuk et al., 2016).

2.11.1 FOSS in healthcare

Free and open-source software is widely used, as there are many advantages to using FOSS systems and there is increasing interest in the use of FOSS within a healthcare setting. It is important to highlight the advantages of such an approach given the focus here on developing such solutions within a healthcare sector.

Brandão et al., (2016) conducted an evaluation of different Business Intelligence software in the healthcare environment. The aim of the study was to compare different systems and provide a recommendation based on several benchmarks. Of the 6 systems reviewed, an open-source system provided the most complete service and scored highest in the ranking. Spago BI was the best tool in the ranking, but another FOSS tool came in second and was so close that it would make very little difference in practice. Ganesh et al., (2016) created an open-source technology to better facilitate healthcare monitoring called The Smart Chair. The paper described the design of a low-cost chair designed for telemedicine and utilising the Internet of Things (IoT). Using sensors and a Bluetooth module it was possible to show information on an Android device which was able to flag any worrying signals in real time. The use of open-source software to create novel solutions to improve the quality of healthcare is apparent in this paper and this has particular significance for this project.

The Cancer Imaging Archive is a free and open source and open-access resource (Clark et al., 2013). This paper described maintaining and operating the archive in relation to the vast number of images it keeps. The paper highlighted the management tasks and user support model from the archive, and the use of FOSS to administer large scale projects within a healthcare environment with a particular focus on cancer. The use of FOSS in such projects allowed for collaboration and the ability to enable better research across multiple teams and venues which would have been difficult to achieve in previous years.

Saeed et al., (2003) presented TM4, which is an open-source data management system for DNA microarray data. The system offered the ability to manage the data and to perform analysis. The system used the MySQL database and specialist data entry platform called MADAM to ensure the complex data was stored correctly. The system had the ability to perform a wide range of analyses including k-means, principal component analysis and self-organising trees. The authors noted that although this was designed to work with a specific type of array data, the system which was freely available to the scientific community could be adapted to work with other types of array data. These papers highlighted the use of FOSS within a healthcare setting and covered many different areas from smart technology working with patients to large scale data management and analysis tools. Brandão et al., (2016) suggested that take up of FOSS systems was scarce at present, but presented a good argument for FOSS solutions as they offered a scalable, adaptable and free option for many institutions.

2.12 Chapter Summary

This chapter has addressed several different topics as this project spans across geography and computing. An overview of the use of GIS measuring access to healthcare facilities was provided before a focus on the use of floating catchment methods. An overview of the initial 2SFCA method was conducted with further details on the E2SFCA method. The different advancements within the FCA methods were described and several tables and graphs showed the timelines and aims of these advancements. This project is focused on access to cancer care and there is an investigation into the utilisation of FCA methods within cancer specific studies as well as a coverage of the use of general GIS approaches in other cancer focused studies that point the way to open-source GIS solution to address variations in access to services. GIS plays a large part in the advancements of the different methods discussed and the development of GIS is described alongside an overview of FOSS systems. A selection of papers shows the use of FOSS in healthcare settings in a variety of different forms. Overall, this review has revealed that a significant focus to date has been on the use of proprietary GIS packages in the implementation of FCA approaches within a number of different thematic areas. However, the increasing trends towards open-source solutions, added to the need to provide user friendly tools that non-GIS experts can use to analyse access to cancer services, suggests that more research is needed to implement such tools in a FOSS environment and provides added impetus for the research conducted as part of this study.

Chapter 3 Proof of Concept Tool

3.1 Introduction

Given the breadth of possible solution pathways that appeared to be on offer with the potential to address the questions raised, it was decided to conduct a pilot to explore some of the possibilities and to narrow down the choice, this allowed the project to focus more effort on pursuing an option that appears to have been qualified as a suitable candidate. To limit the time investment, it was decided to simplify the task to adopt only a straight-line distance solution to feed into FCA computations. The findings from this stage of the research guided the overall approach taken in the project.

This chapter explores in detail the desktop tool that was used to calculate straight line 2SFCA analysis within a standalone Geographical Information System (GIS). The strengths and limitations of using such systems to perform these calculations are addressed. Following a review of the options available to address the objectives of the study, and the rationale for the choices made, a description of the potential user and the nature of the user experience was investigated. The data used was critically reviewed and explained prior to a review of the tools' performance across several different types of data and task. Visualisation outputs were utilised to display the results of the testing and these are described and interpreted. This initial design was used as a concept; similar programs to this have been developed using proprietary software (Higgs et al., 2018; Langford, Higgs and Radcliffe, 2018) but there are no suitable FOSS alternatives. This tool has been designed as a proof of concept to judge the suitability of this approach and to investigate a range of the options available. It has been used to better understand how the 2SFCA algorithm can be translated into code and to highlight some of the problems that will be need to be addressed whether they be computational, data led or usability focused. Through the development of this tool, it was possible to get early user feedback from researchers within Tenovus that could enhance further iterations.

The software described in this chapter has been created in QGIS which is an open-source GIS with wide ranging functionality ranging from simple mapping to creating videos and spatial analysis (QGIS 2 Cookbook - Mandel, Ferrero, Graser, Bruy, 2016). QGIS provides a tool to build plugins called the Plugin Builder which was used for this proof-of-concept. Using Python, it was possible to use the functionality of QGIS in order to manipulate data to perform different tasks in the form of a plugin, which enabled the user to utilise the functionality of QGIS whilst manipulating data (QGIS Python Programming Cookbook, Lawhead, Google Books, 2015).

This aim of this project was to develop solutions around the use of the 2-step floating catchment area (2SFCA) technique, in order to better understand accessibility to cancer services in Wales. These models incorporated a supply to demand ratio which included both distance to a service and the population that could access that service. It has been shown to give a more realistic understanding (see section 2.1) of the services available than more simplistic measures of accessibility that are commonly used for analysis (Apparicio et al., 2008). The plugin tool uses only straight-line distances to measure the distance between the supply point and the demand point, prior to a fuller consideration of more advanced methods for calculating the distance between two points, including creating road networks with weighted speeds (see section 4). Previous studies included advancements to the original 2SFCA algorithm (see section 2.3) that included the incorporation of distance decay functions, which assigned weightings to supply and demand values depending on the distances between the respective points of provision and population base. By utilising the so-called enhanced 2-step floating catchment area (E2SFCA) method that included distance decay functions, it was possible to get more realistic results than those based on the original 2SFCA algorithm (Luo and Qi, 2009).

The performance of the tool was tested using several data sets; the purpose of the testing was to see how well the tool coped with large amounts of data and the times taken to complete these tasks. The testing showed that there was a linear pattern where the more data that is input, the longer time is taken for computation. This tool performed well with smaller data sets but did not scale appropriately as the data increased (see section 3.9). The tool loops through the data, this means that many more iterations through the data is necessary to get a result and without the use of a spatial database it would have been very difficult to perform the calculations on these datasets in an acceptable time. This was due to the ability of the spatial database to use spatial indexing, and bounding boxes to remove a large amount of the computation.

The prototype tool was critically evaluated to understand what has worked and what, if any, are the limitations. Using a plugin in this environment offered many benefits compared to writing a standalone piece of software, including the development time and the ease of visualisation. There were also limitations to what the plugin could do. It was difficult to create a routable network in this environment as access to the network analyst feature was difficult, and without the use of a spatial database it was impractical when using large data sets; the times taken to compute moderate data sets without being routed were very large as shown in table 3.1 (p. 76).

There were multiple data sets used in this section, including census data, cancer data, and the Welsh Index of Multiple Deprivation, derived from numerous online sources. This drew on the most detailed data sources available within current ethical guidelines. Importantly, privacy concerns were addressed by discussing the data required with Tenovus and agreeing the best way to abstract the data to be within the current General Data Protection Regulation (GDPR) guidelines.

Visualisation outputs can be used to provide a wider understanding of the FCA approach; in this instance, maps were used to display the results in a way which was easily read and understood. When communicating complex geographical data with non-experts it is preferable to use map-based displays for the output of the accessibility study as these are likely to be more readily understood than alternative statistical and tabular output.

This led to an action plan, helping to identify which aspects of the prototype were successful and could be adopted in the main study and which areas needed to be addressed such as the adoption of alternative solutions to the distance calculations. These findings were used to guide the project and ensure that most of the time was focused in the right direction and minimal time was wasted on unsuitable solutions.

3.2 Open source

One of the fundamental principles of this project was to create an open-source tool that is able to benefit a wider range of users. This required the use of open-source solutions that were fit for purpose. The Open-Source Initiative described open-source software as “software that can be freely accessed, used, changed, and shared (in modified or unmodified form) by anyone. Open-source software is made by many people, and distributed under licenses that comply with the Open-Source Definition.” (The Open-Source Definition | Open-Source Initiative, 2007).

There were advantages and limitations to using open-source software and this needed serious consideration with each decision on how to move the project forward. The main advantages of using open-source software were that it is free to use, which can save companies large sums of money compared to what they would pay for proprietary software licences. It is also continuously being improved by those that are developing the software, and it is possible to tailor the software to specific individual or organisational requirements by using in house developers or hiring an external company. By working in an open-source community, each contribution can add to the whole, and helps to foster an innovative culture where all of those involved share information for greater knowledge.

The main issues with using open-source software are that although the software is free it may cost money to keep the software current and working to the needs of the user (development of specific tools, additional training costs, more specialist help when problems arise). There is also a risk that the key developers updating and maintaining the software move onto different projects and the development will stop. There are issues with some software where documentation is not being as well written as that of proprietary software, and it is important that the user understands this before moving to open-source software. Security is another potential area of concern and it is important to clarify that as the user has access to the source code so does any potential malicious user who could try and exploit any potential weakness in the way that the program is written. This needs to be understood by potential users, and organisations need to ensure they address any potential security risks.

Most of the programs discussed in this chapter are part of OSGeo, which is the open-source geospatial foundation[9]. They have various spatial projects that they promote under their banner, including QGIS, PostGIS, GRASS GIS and OpenLayers plus many more. By using large communities like this, it is possible to mitigate the risk of the projects failing as there are many dedicated contributors who are passionate about creating better open-source software. Most of the tools are used by very large organisations and are able to rival the proprietary solutions in terms of functionality.

The benefits of using open-source software outweighed the negatives for this project, but it was important to highlight the issues. By being able to create a free and easy to share tool which is able to deliver good quality healthcare accessibility calculations, such an approach has the potential to enhance the work of organisations concerned with the planning of healthcare services who may not have the finance or resource for more costly alternative solutions.

[9] https://www.osgeo.org/

3.3 QGIS

QGIS is a free and open-source desktop GIS, which supports multiple platforms (Windows, Linux, OS etc.). It integrates with many additional packages such as PostGIS, GRASS, etc. and allows the user to install plugins to provide numerous tools for the user giving the user the ability to tailor the package to their needs and specialities. QGIS has been available since the late 2000s and is being used more regularly in industry; it is one of the main FOSS alternatives to ArcGIS^{TM} (see a list of contributors at (Sustaining members and Donors, 2020)), a proprietary GIS that is the world leader in GIS software sales. There are alternatives to QGIS; as well as ArcGIS: SAGA GIS, GRASS GIS, gvSIG, SuperGIS, Mapserver and Google Maps Engine are just a few of the GIS systems currently available, and these all have their specialities and strengths in different tasks.

QGIS best fitted the requirements of this project at this stage because it was easy to learn (multiple courses offered online and at various venues) and provided the capability for plugins to be created. One of the primary aims of this pilot study was to demonstrate that the creation of an easy-to-use desktop 2SFCA tool was feasible; QGIS provided an ideal environment for this to be tested within due to its current stage of development, as well as the large community of developers associated with the project which ensures a stable and reliable base platform. QGIS also provided the ability to create and modify shape files, which was the primary format for the data used in this study. There were various other ways that the data could be stored and QGIS allowed data to be transferred from GeoJSON, KML, and most of the other formats used to store geospatial information.

Although there are benefits to using QGIS, it also has some downsides too. For example, it is not as complete as ArcGIS, which offers far more built-in functionality such as a built-in spatial database and easy to use network analyst tools. Google Maps have some innovations which QGIS is not able to replicate at this time such as its routing capabilities utilising the prevailing road conditions to give a more accurate time or route and the more up to date satellite imagery. But, even acknowledging these drawbacks QGIS is being used by various organisations across the world as an alternative to proprietary GIS packages as a result of the savings being made in implementing the software, the potential for customisation, and continuing advancements in visualisation and analytical capabilities. There are also a variety of different paid and free training programs offered from a wide range of sources to educate users in the use of QGIS at any level.

QGIS allows users to create their own plugins using the plugin builder tool; this uses Python scripting and PyQT creator to include functions that are specifically desirable for particular users or groups. Using the plugin builder, for example, it is relatively straightforward to create the 2SFCA tool, which allows the user to load data sets directly into QGIS, and then manipulate the data by performing the 2SFCA calculations. Once the calculations are completed, it is then possible to show the data in QGIS using the built-in mapping functionality. QGIS also amends the shape file to store the results that can be viewed in a tabular format.

Figure 3.1 shows the QGIS environment (version 2.18) familiar to most users of windows machines that follow a similar format to most of the Microsoft Office suite and which is highly customisable by the user. The main menus are positioned along the top with many buttons underneath which are interchangeable depending on the workflow so that the most used tools are easy to access. The rest of the screen mainly consists of a large viewing pane and several boxes that are used to select and manipulate the data contained within it. The large purple map is one layer of data that shows the LSOA outlines in Wales, and each yellow dot is the population-weighted centroid for each LSOA.

Figure 3.1 shows the QGIS environment (version 2.18) familiar to most users of windows machines that follows a similar format to most of the Microsoft Office suite and which is highly customisable by the user.

Figure 3-1 QGIS 2.18 Layout

3.4 Python

Python 2.7 is a high-level computer programming language which was used to create the prototype tool. Python was initially released in the early 1990s and it has become increasingly popular in the last 10 years, with major corporations such as Yahoo, Google, IBM and NASA now using it. The main reason for the increased usage is that it is highly adaptable, offers fast and efficient processing, and is supported by a wide range of libraries to add extra functionality, which makes it a very fast and functional programming language when compared to Java and C++. programs written in Python normally require less lines of code when compared with many of its the main rivals, and the way in which the code is written makes it easy for other developers to read and understand.

There are multiple benefits to using Python for the developer and the end user. Because Python is so readable, it makes it much easier for legacy code to be maintained by different developers. It also works across multiple platforms and is not tied to any specific environment unlike other languages such as C#. Although it is easy to use it is still a very powerful object-orientated language that allows the developer to create complex programs. The libraries available for Python are varied and of good quality, and there are a wide variety of specific libraries which can be used to make complex problems easier. NumPY and SciPY are particularly important widely used libraries that contain formulas for completing common mathematical and scientific calculations without having to write code to perform these calculations. The use of such libraries helps to make Python code easy to read yet very effective.

Python has its imperfections and some tasks would be completed more efficiently using a different language. For example, it is an interpreted language, which means that it is not compiled like C++ or Java, and therefore can take longer for the code to execute. Also, the way that memory is managed in Python makes it less efficient than Java or C++ for tasks which involve high levels of memory management. These downfalls do not stop it being a good choice for this project, and since it was the only language that could have been used at this point with the plugin builder the choices were limited.

3.5 PyQT Creator

PyQT Creator is a tool used to create the interfaces used by the plugins. It enables the developer to use QT creator by porting the Python language into QT. QT is normally used with C++ and it is a very mature cross-platform user interface creator. It makes creating an interface simple and helps the developer to visualise what the interface will look like. By using a drag and drop system it is simple to make interfaces that do what is required for the project, and these are easily changed and manipulated when amendments are necessary.

There are a range of different interface creators available but QT creator is one of the main cross platform choices and is used commonly by developers (Bhalla, 2008; Binder et al., 2017). Microsoft has Visual Studio that performs a similar task, but it is not used to create cross platform tools as often. Wxwidgets, Gtk and Fltk are all cross-platform alternatives but QT is used in this stack and has been shown to work well (Rischapter, 2014).

Figure 3-2 shows the interface that was created. The tool takes supply and demand inputs that the user specifies and applies them to a catchment size that needs to be adjusted by the user. Finally, the user is able to specify where the tool outputs the results. There is some management information recorded and this shows how many supply and demand point pairs there are and how many pairs there are within the specified distance as well as a start and a finish time when the calculation is performed. Once the results have been completed, they need to be joined with a separate layer so that the data can be displayed in visual format. For this analysis a generalised LSOA layer has been used.

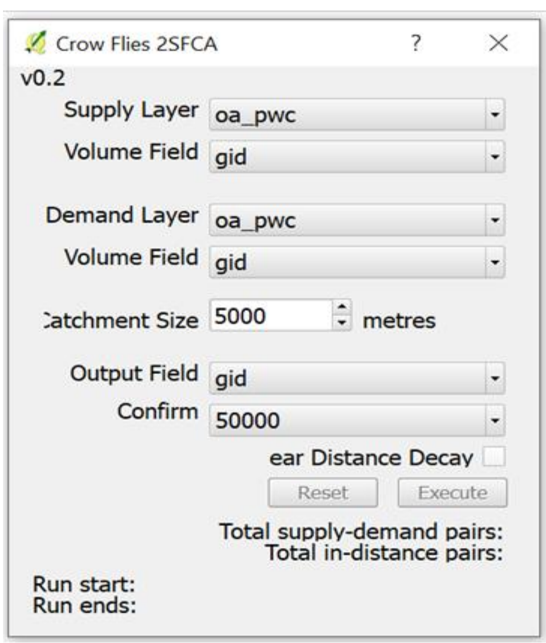


Figure 3-2 Crow Flies 2SFCA tool

3.6 Straight Line 2SFCA

To compute 2SFCA scores, the required calculations needed to be converted into a series of commands that could be coded into the program. Figure 3-3 shows the basic steps of the algorithm; the program loops through each supply then demand point and then adds the attribute information of those with the specified catchment area into a dictionary. The second step loops through each demand then supply point and adds the attribute information to the dictionary. The two are summed to give the 2SFCA score of each supply point which accounts for how close a demand point is to a supply point, and also for how many demand points can access the supply point.

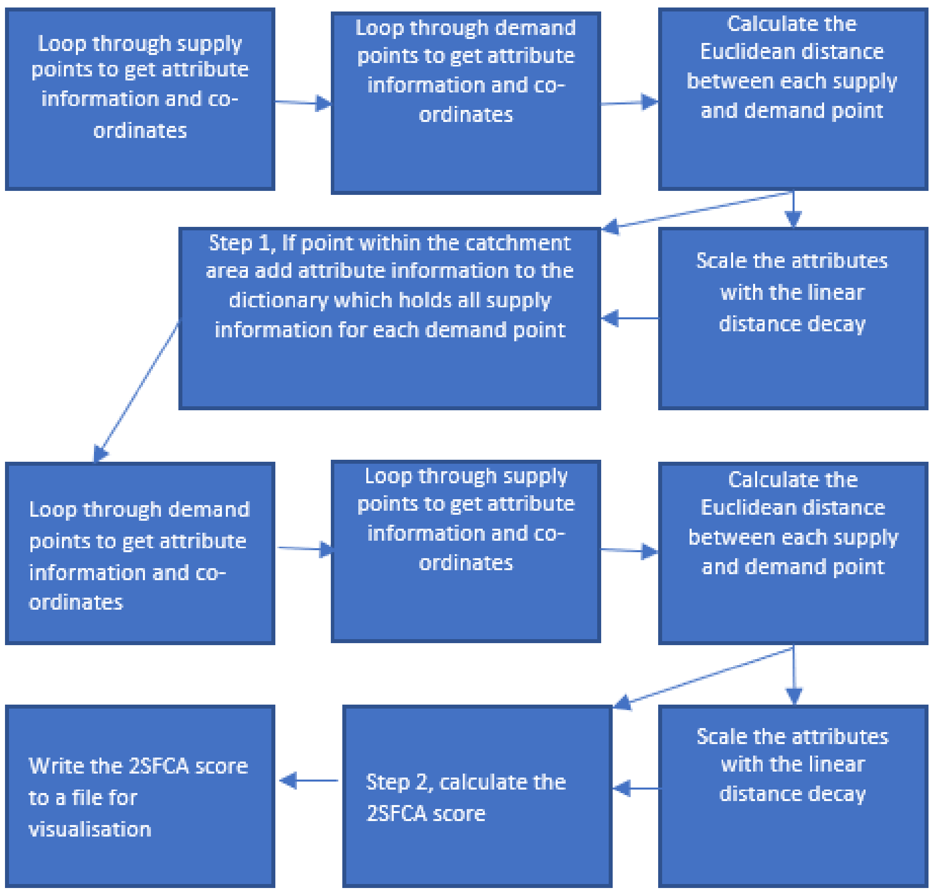


Figure 3-3 Algorithm flow chart

To loop through each supply and demand point is a very simple way of collating all of the information, but it creates problems when using large datasets, as even on modern computers it will take considerable time to loop through the multiplicative combinations. The way this program was written also means that it must loop through the points multiple times. This could possibly have been improved upon by saving the information of intermediate stages to file storage, although this too can be a time-consuming process.

This tool uses straight-line distances for the purpose of the calculations and this is not the most effective way to measure the 'geographical friction' between one place and another. It would be preferable to use the time taken to travel between point A and point B or the true network distance. To get this information it was possible to create a routable network that used information pertaining to all the roads and connectivities between point A and point B, and calculated the time taken to travel via the shortest network route between them. Using different graph traversal calculations, it was possible to get a rich understanding of the time taken and distance travelled. It may be more feasible for someone to travel ten miles on a motorway rather than five miles on small B roads, for example

It was possible to create networks that were more detailed than just vehicle travel times and they could be expanded to include walking, cycling and different forms of public transport that further added to the accuracy of the catchment sizes. There were some problems with using a network to calculate the catchment sizes as they needed to be kept up to date and this can be a time-consuming task. If a new road is built or a road is closed then it can have a severe effect on the catchment size and this must be considered.

A linear distance decay function was coded into the tool to help improve the accuracy of the results. This function assigned weighting values to the points within a catchment area based on their distance from the target, with a weighting of 1.0 being the closest and 0.0 being the furthest away. All points in-between got a corresponding value which declined the further away the point was. By using a distance decay function the results were more accurate as it took into account the fact that someone who lives next door to a supply point has better accessibility than someone who lives at the extremities of the catchment area.

There are various distance decay functions that can be used, and many have been tested in healthcare accessibility, with the advancements of each being discussed in the literature review (sections 2.3 and 2.4). There is no right or wrong function to be used but depending on the data, some have shown to give more intuitive results than others, Dai, (2011) used a Gaussian based distance decay function and Luo, (2014) used a continuous distance decay function. Many authors (sections 2.3 and 2.4) suggest that the results that show what they would expect are more intuitive, for example, the 2SFCA method would have less refinement than that of an E2SFCA method, it does not account for any movement within the catchments and may not be as good at describing the data, if the results showed high access in areas with no supply then this would be unintuitive and would need further investigation to make sure there are no errors with the method being employed. It is important to understand the data that the prototype tool uses in order to be able to tailor the distance decay functions to the data. In the software developed in the main phase of this study several alternative distance-decay functions were programmed into the system so that the user is able to experiment with the effects these settings may have on outputs computed with the data at their disposal (see section 4.2.4).

3.7 Data

To complete this proof-of-concept study a number data sets were used. For the tool to work the spatial data first needed to be stored in a readable format such as shapefile or GML. When the data was in the correct format it was then easy to manipulate and visualise using QGIS once again. Experiments with the prototype used 2011 census data to report population demand totals and data supplied by Tenovus to report the locations of supply side sites. Furthermore, data was used that reported the Welsh index of multiple deprivation scores and hospital data.

Lower Layer Super Output Area (LSOA) is a type of geographical unit used in England and Wales[10] typically to display census data at a local level. Output Areas (OA) are constituent units used to build LSOAs, and the smallest geographical unit for which census data is released. The OA and LSOA on average are composed of around 125 and 650 households respectively; the areas tend to be smaller in areas of high population and larger in rural areas. In Wales, at the time of the 2011 Census of Population, there were 1,909 LSOAs and 10,036 OAs. They are used in this paper as they are the smallest geographical units for which much of the data is released.

Population-weighted centroids were used to represent a single geographical location for each LSOA and OA; they are single points falling within areas that represent the distribution of the

[10] https://www.ons.gov.uk/census/2011census

population in each area at the time of the last census. These are commonly used in this type of study as they are easy to understand and can be used to represent where the population is assumed to be located. Total population was also used, taken from the 2011 census, to provide an estimate of the level of demand present in each LSOA.

Mobile unit data was provided by Tenovus and consisted of the locations used by the three mobile units but not the time spent at each location. Thus, an arbitrary supply value of one was used for this data. There were 27 sites in total that were visited by the three mobile units and all of these were considered to be of equal value for the purposes of the analysis presented in this chapter. The hospital data has been put together by using websites to try to understand the spatial pattern of 'static' cancer provision in Wales.

The Welsh Index of Multiple Deprivation combines various different metrics to present an index score to describe levels of deprivation in an LSOA area. One of the metrics used in the construction of the health domain is cancer incidence (indirectly age-sex standardised) (number by 100,000) at LSOA level and this figure was been used to create a demand estimate (to compare with the use of the total population derived from the 2011 census).

3.8 Locations of cancer facilities supply and demand

Figure 3-4 shows the sites used to locate the mobile units; as stated before there are 27 sites in total and the spread is mostly in the areas with higher population around the south east and the coastal regions. These are sites where Tenovus is able to locate its services and they are normally decided upon by the health board. In addition, local site factors such as parking availability and accessibility restrictions are important considerations that Tenovus needs to take into account in the final choice of sites.

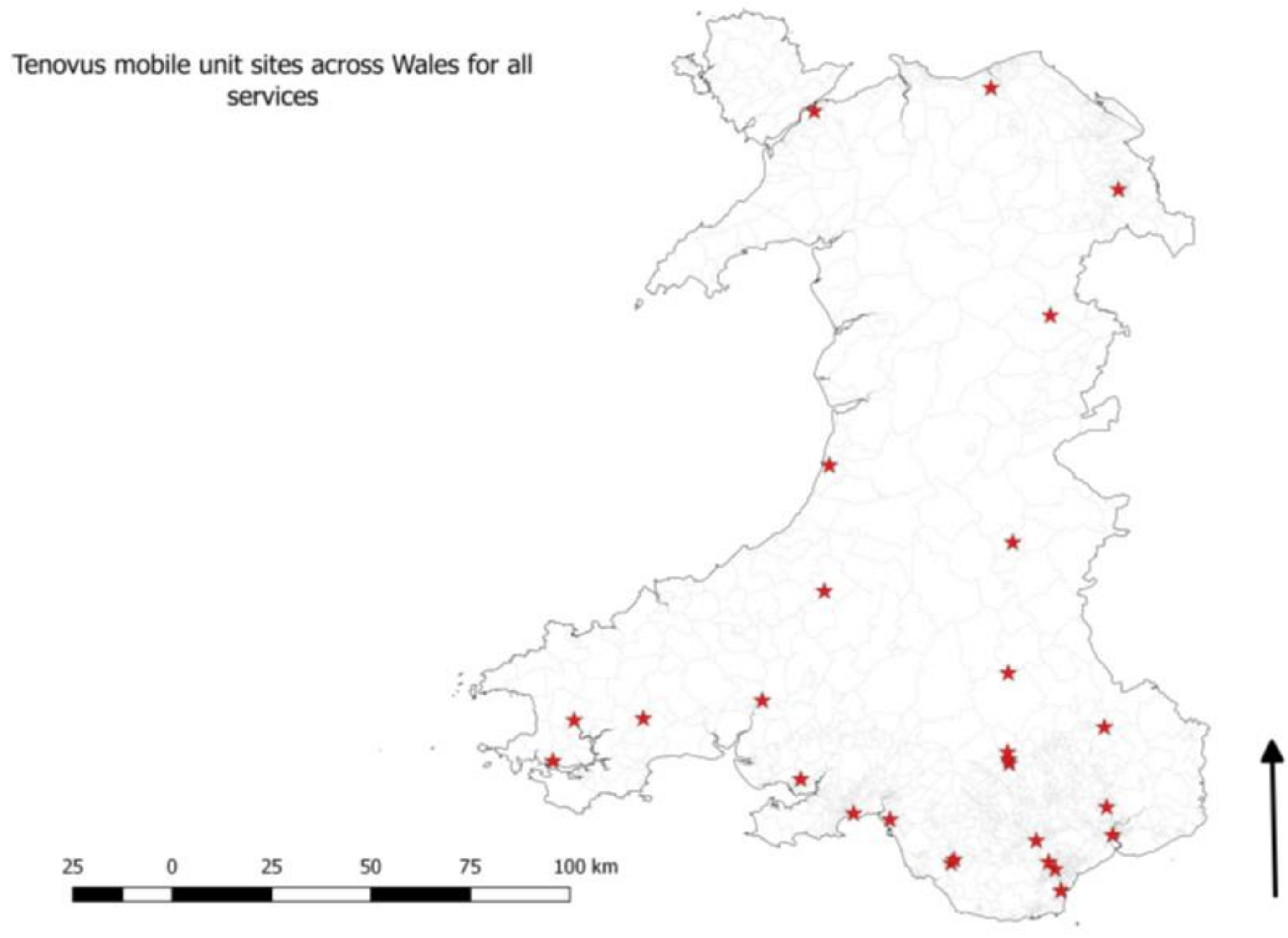


Figure 3-4 Tenovus Mobile Unit Sites across Wales

The maps of 2SFCA scores computed by the prototype tool display some interesting patterns. When computed using smaller catchment sizes, it was possible to see a ring effect around the van sites. This arose because the surrounding LSOAs had good accessibility, which provided a high accessibility score, and this clearly declined with increasing distance. It is important to note that although Cardiff and the south east of Wales had the highest concentration of supply sites, they did not register the best levels of accessibility. This was due to the fact that there were more people and a lot more LSOAs in the area. This had the effect of spreading the service supply much more thinly than corresponding locations in mid Wales. As a consequence, it was the latter which registered the higher accessibility scores as there was much less demand for the same service from within the local area.

When the modelling was performed using small catchment sizes there were a large number of areas that reported no accessibility at all, which meant that if say, 10,000 metres was stated as the furthest distance an individual would be prepared to travel for these services, then the locations of the vehicles needed to be spread much more widely throughout Wales to achieve a good and uniform coverage particularly in rural areas.

At a more reasonable 30,000 metres (<19 miles) almost everyone on the map had some level of accessibility and with one or two strategic van placements everyone in Wales would have some degree of accessibility even if it was not ideal. This continued as the map got to 40,000 metres and most of the map had some accessibility to the services.

This data did not take into account any other mobile cancer services, which would have been helpful in understanding the effect that the Tenovus sites had on general accessibility to cancer services in Wales.

Figure 3-5 shows variations in cancer incidence (indirectly age-sex standardised) (number per 100,000) at LSOA level across Wales with areas with higher levels of cancer incidence in darker shading. Such variations were used to estimate the potential demand for cancer services which could be incorporated into the final FCA calculations.

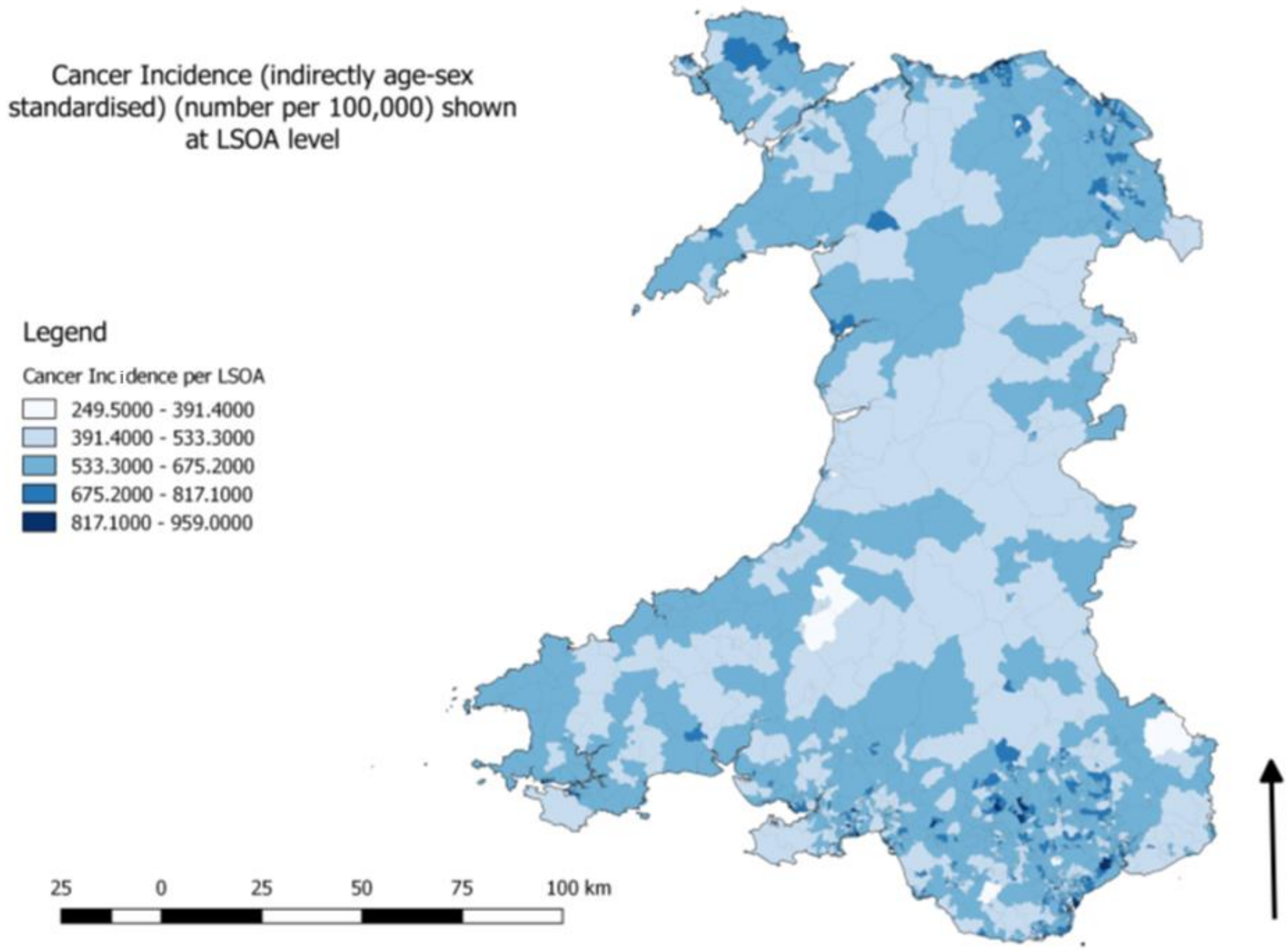


Figure 3-5 Cancer Incidence (indirectly age-sex standardised) (number per 100,000) shown at LSOA level

Figure 3-6 shows all of the LSOAs with their total population. It is possible to see that the LSOAs vary in size with the larger LSOAs in mid Wales, which has smaller populations compared to the very small LSOAs in and around the Cardiff Area.

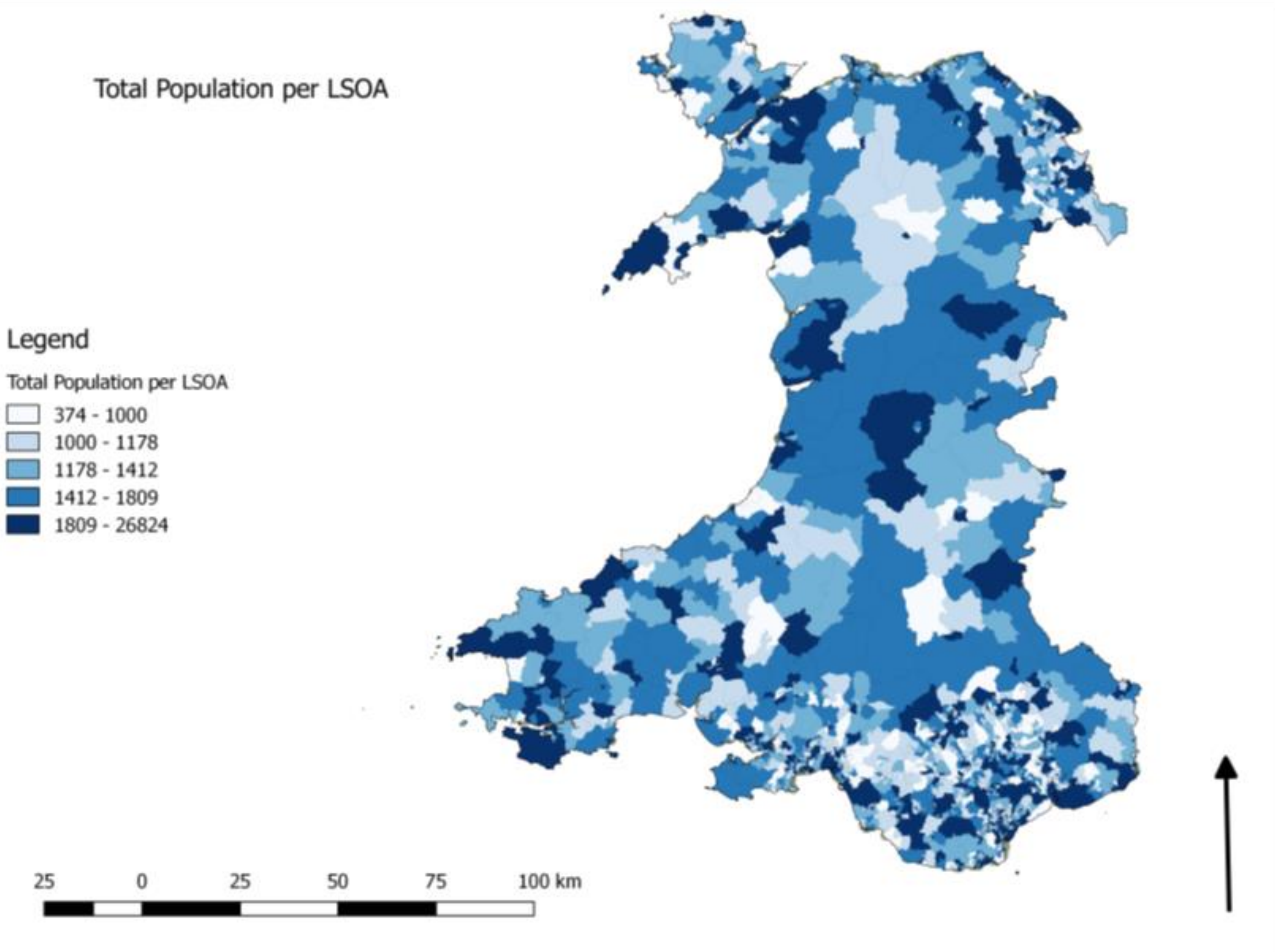


Figure 3-6 Total Population per LSOA

3.9 FCA visualisations

Using the data shown in section 3.8 it was possible to use the proof-of-concept tool to run several FCA analyses. This was completed by entering the supply and demand catchment layers into the plug assigning supply and demand fields then setting a catchment size. The results were then joined to an LSOA layer which could be styled to show the results using different colours. For the maps in this section a quantile function was used to categorise the results into bands which could then be visualised.

Figure 3-7 shows that there are areas in Wales with poor access to Tenovus mobile units within ten kilometres of their homes. As can be seen there is unsurprisingly a high level of accessibility in the LSOAs that have the mobile units visit them, but then there is no accessibility in those areas surrounding this site. It is possible to see that in Cardiff and Swansea even though these areas have the largest amount of supply, because of the way in which the FCA scores are calculated, they do not necessarily have the best levels of accessibility due to the large volumes of people that are potentially accessing the services.

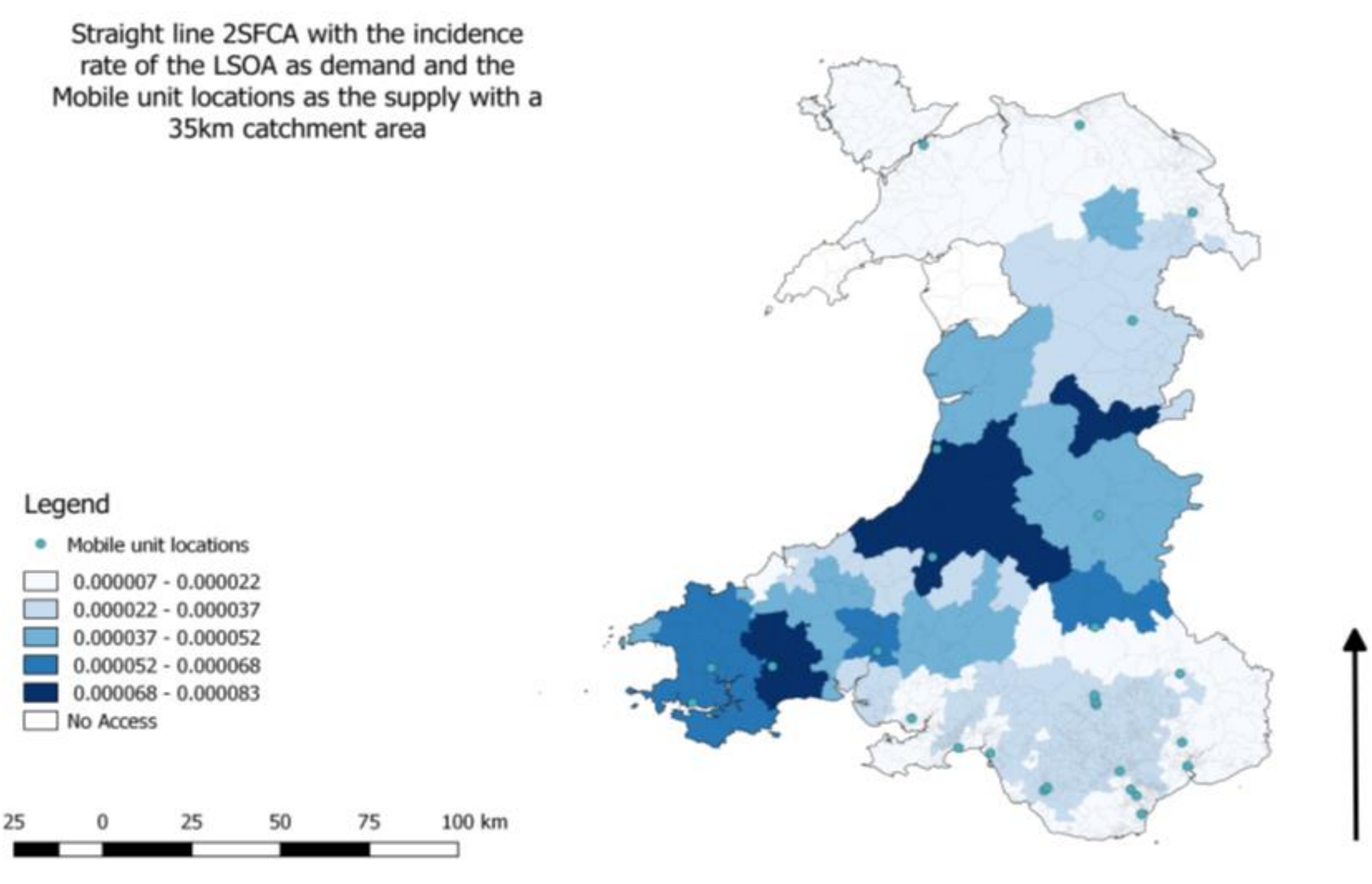


Figure 3-7 35 km Tenovus Vehicle Sites as Supply LSOA Cancer Incidence (indirectly age-sex standardised) as demand.

Figure 3-8 highlights the effect of the catchment size; as the size of the catchment area increases there is a general smoothing effect. This shows that as more people have access to the service it becomes less accessible and if this were to continue until the catchment size was as big as Wales then it would result in all areas having the same (rather low) score.

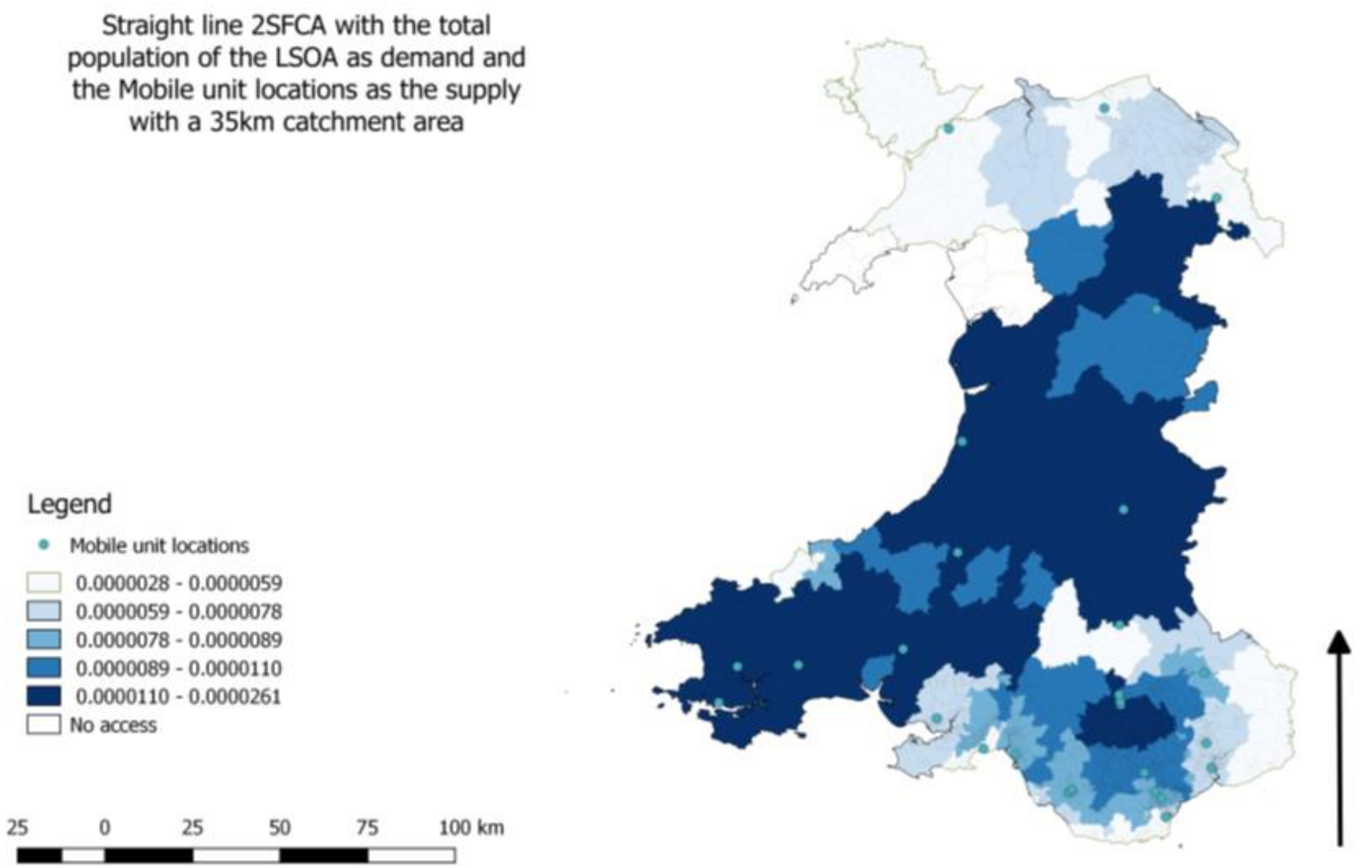


Figure 3-8 35 km Tenovus Vehicle Sites as Supply LSOA Total Population as demand.

Figure 3-9 shows very similar patterns to figure 3.8, which suggests that cancer incidence and total population are linked. The spreading is still occurring and there are definite spikes in accessibility depending upon the location, and there is a more even level of accessibility in the Cardiff area.

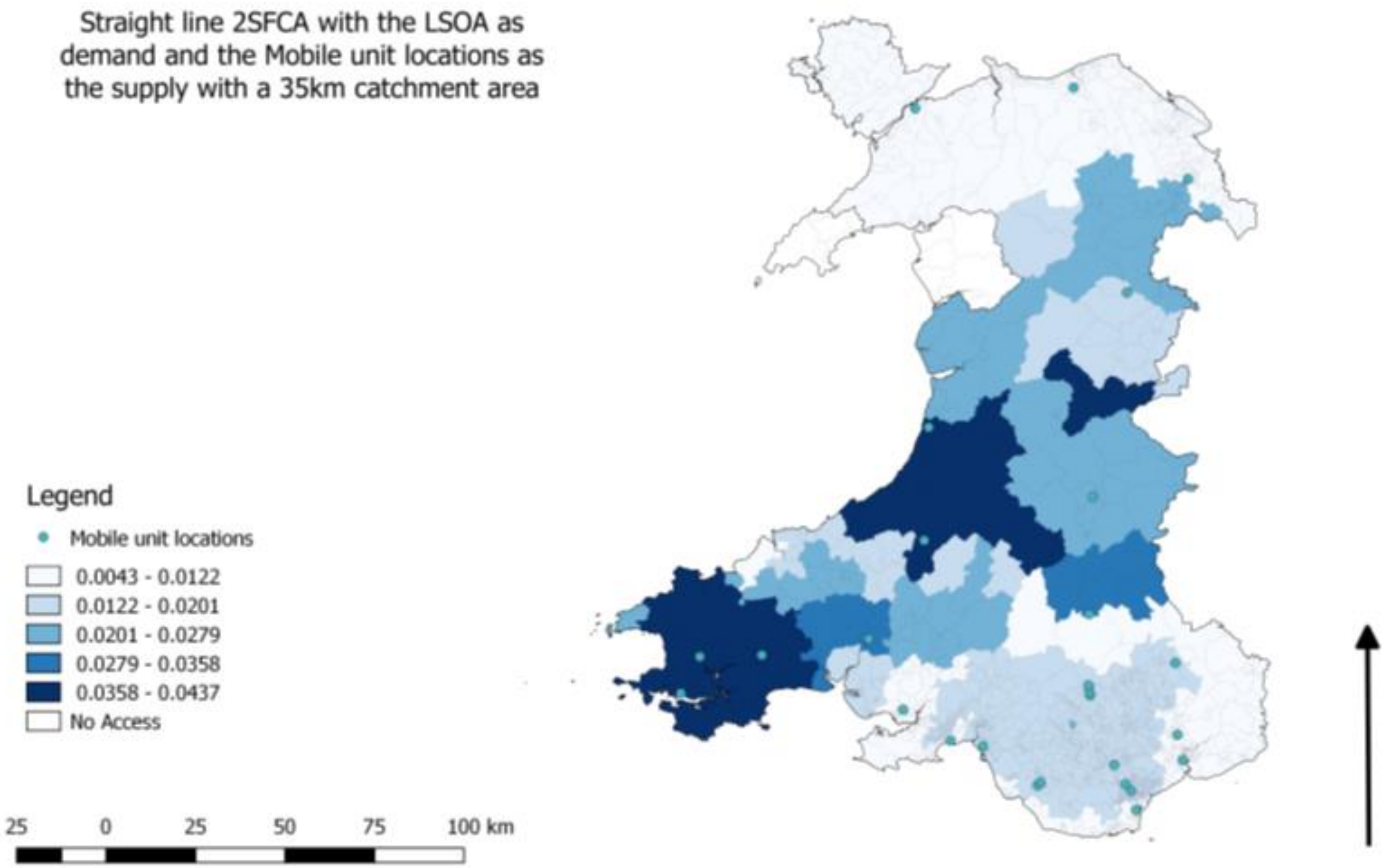


Figure 3-9 2:35 km Tenovus Vehicle Sites as Supply and LSOA as demand.

Figures 3-10 and 3-11 show all of the static hospital sites in Wales that cover a variety of healthcare needs and the second map shows all of the hospitals that have staff dedicated to cancer care. On the cancer services map there is a score next to each point, which is the supply number given to each point and this number is based on the information available from the health boards' websites.

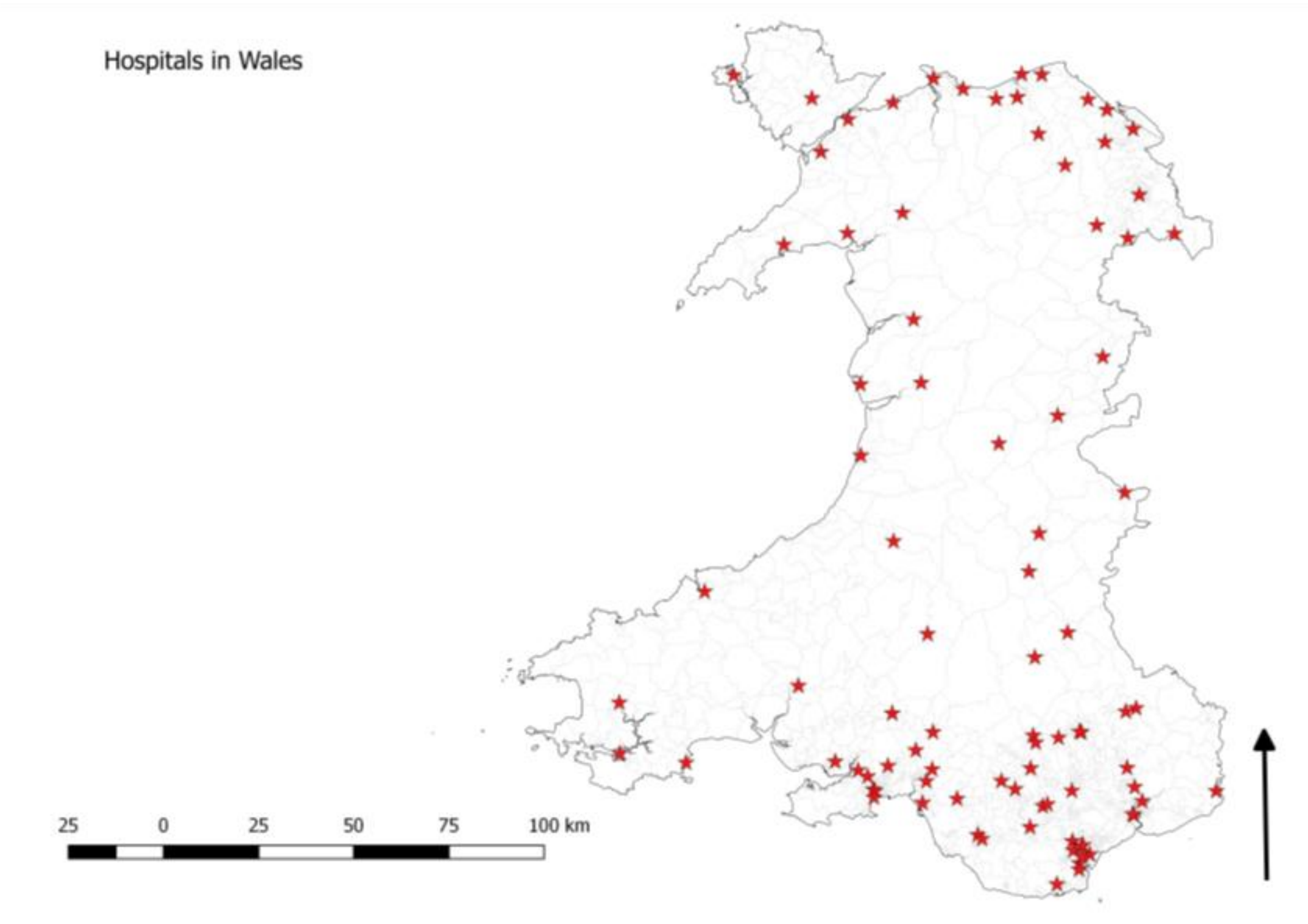


Figure 3-10 Static Hospitals in Wales

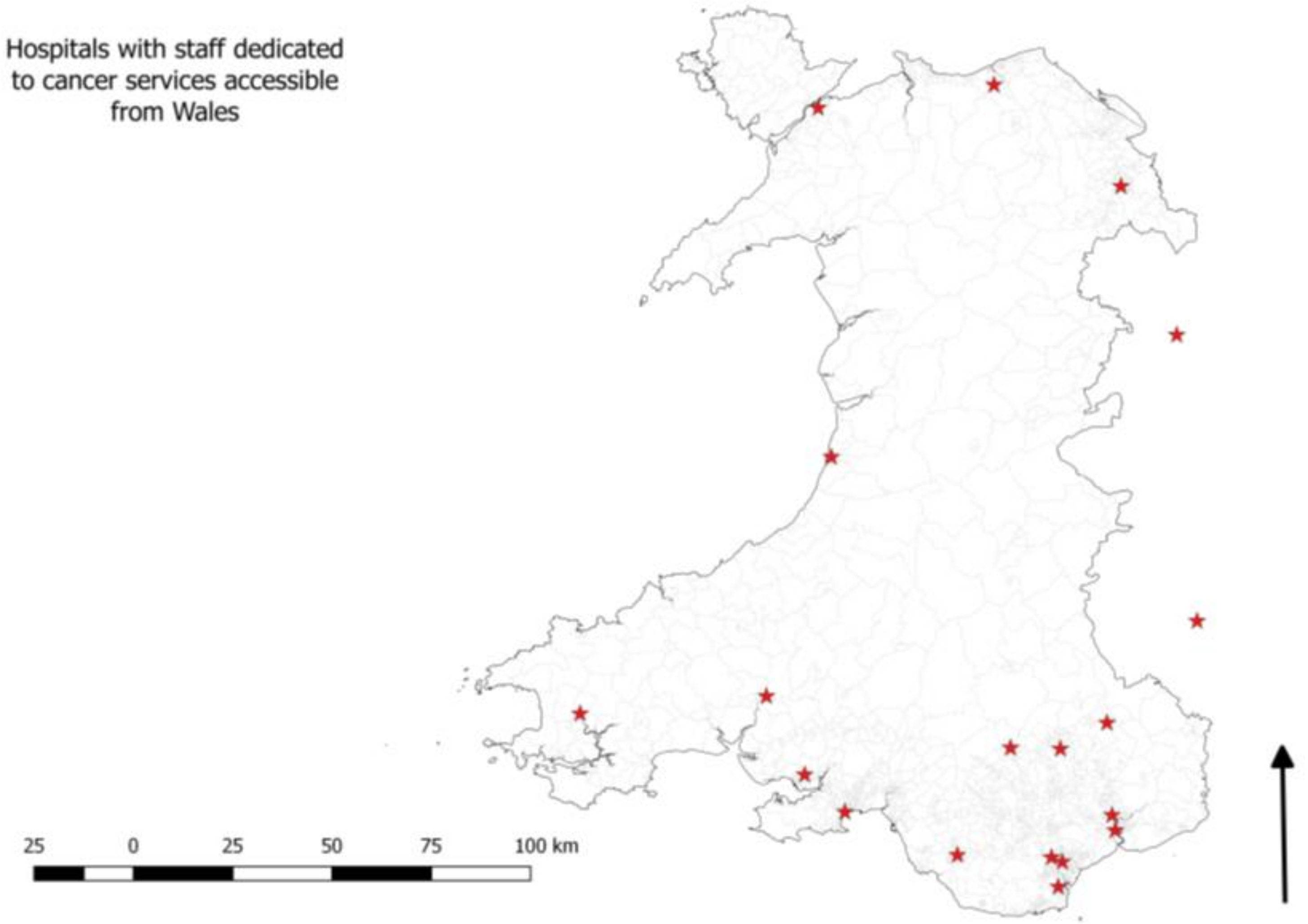


Figure 3-11: Hospitals with dedicated cancer staff accessible from Wales

Figure 3-12 shows that there is a relatively uniform level of accessibility across Wales with a couple of spike points across the country. The southeast and northeast would appear to have slightly higher levels of accessibility than the rest of the nation and this would fit in with these being areas with larger populations than mid and west Wales.

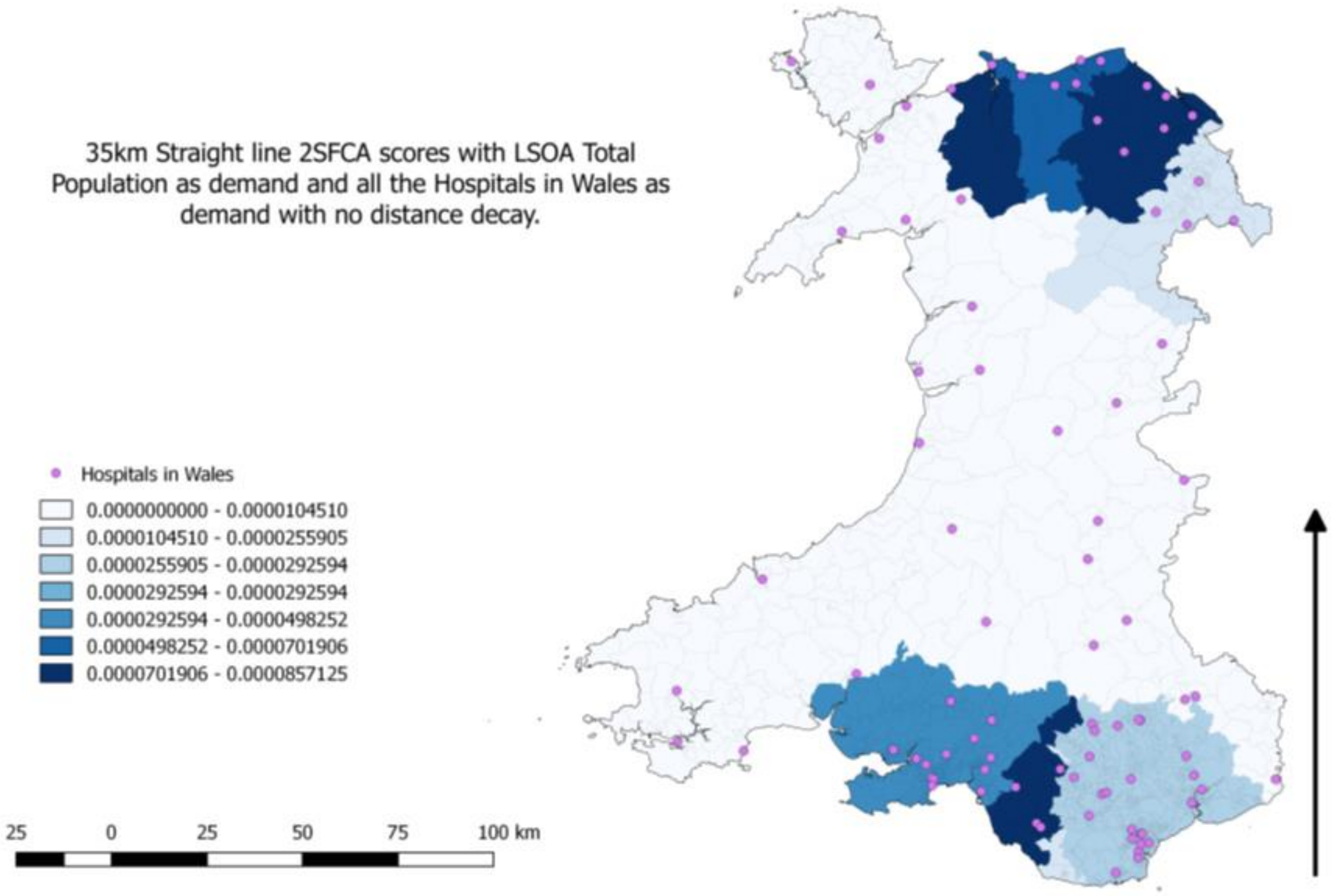


Figure 3-12 35,000- metre Hospital Sites as Supply LSOA Total Population as demand.

When figure 3-13 was compared to the Total Population maps, there was a clear difference. These maps showed a distinct difference between the southeast and the majority of the rest of Wales. There was also a spike in accessibility in the north west of the country as well. This suggested that although there are more people in the large cities and that this is where most of the hospitals are based, there were potential inequalities in access to cancer services that were worthy of further investigation.

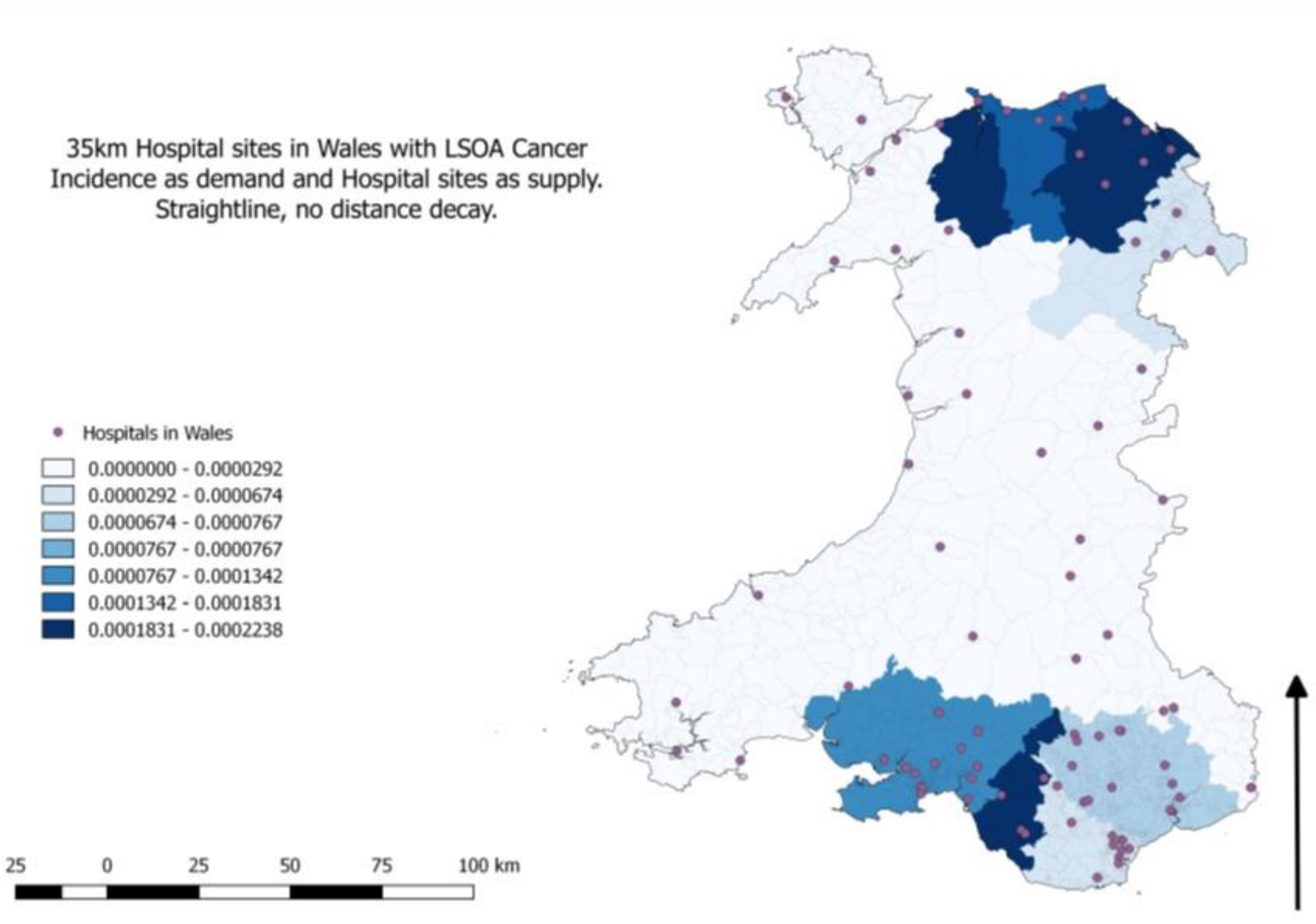


Figure 3-13 35,000- metre Static Hospital Sites as Supply LSOA Cancer Incidence (indirectly age-sex standardised) as demand.

As there were very few hospitals offering specialist cancer care, there was a clear and distinct pattern in these maps that showed that most of the resources for cancer were based between three sites across Wales. With Singleton Hospital, Velindre Cancer Centre and the North Wales Cancer Treatment Centre supporting the northern and southern areas of the Country (Figure 3-13). This showed that anyone living in west or mid Wales had to travel a large distance to gain access to the appropriate levels of care that they needed. These maps were not as detailed as they could be, but proved to be a useful tool when communicating with Tenovus to show the type of visualisations which could be expected in the next iteration.

3.10 Performance of Concept Tool

The charts below (figure 3-14 and 3-15) show the relationship between time and catchment area sizes. This varied as the dataset changed, but it was clear that the larger the dataset the more time is taken to complete the calculations, and the larger the number of pairs that need to be printed to the attribute table. Although the calculations took the majority of the time to compute, and at the moment, they were all being computed for every pair, it would still stand to reason that the larger the amount of data that needs to be reported back the longer the process takes.

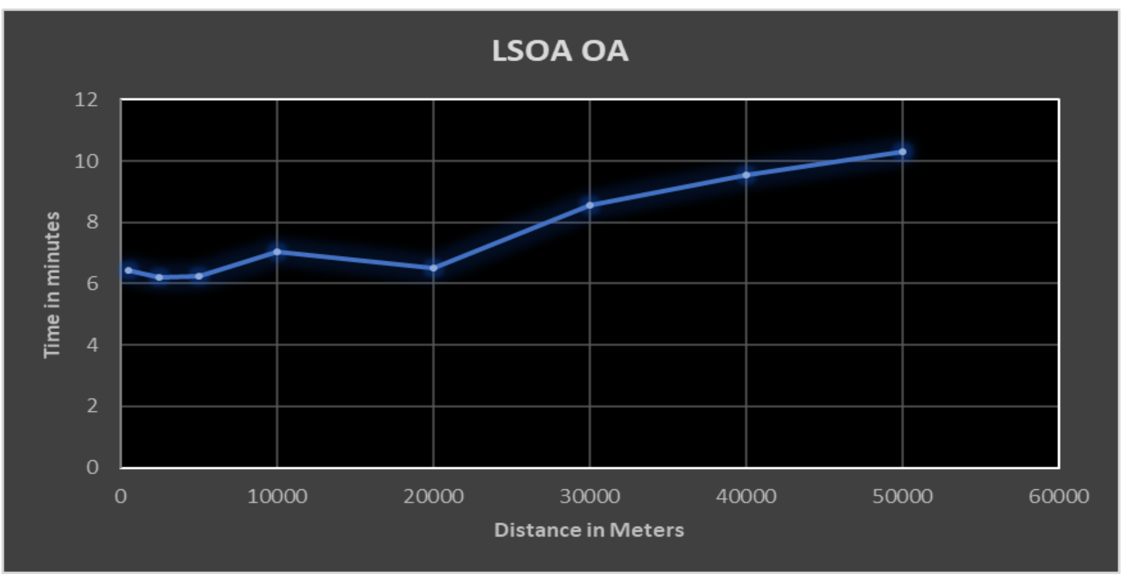


Figure 3-14 Graph showing times compared with catchment size; x axis is the time in minutes, and the y axis is the size of the catchment area in metres

The following graph shows how the number of in catchment pairs rises as the catchment size increases. This is data dependent, but would normally trend upwards as it would be more likely that the larger the catchment size the more points fall within it.

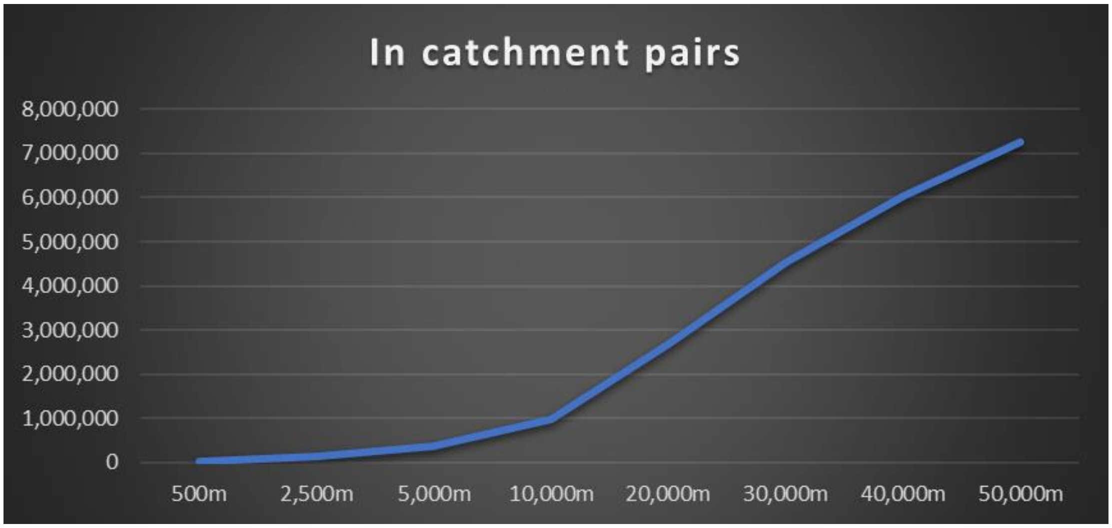


Figure 3-15 Graph showing in catchment pairs compared with catchment size where the x axis is the number of pairs calculated, and the y axis is the size of the catchment in metres

Table 3.1 shows in more detail the full results of the test run and how many in distance pairs were computed. This was important to know as the exponential rise is shown above in the time taken to complete.

Distance	**Time**	**In catchment pairs**
500m	6.45	14,453
2,500m	6.20	145,334
5,000m	6.25	382,935
10,000m	7.04	992,564
20,000m	6.50	2,690,238
30,000m	8.57	4,551,756
40,000m	9.53	6,030,083
50,000m	10.32	7,249,942

Table 3.1 Table showing distance, time and the number of in-distance pairs

As a further example to show the exponential rise in time, a test was completed at a 50 km catchment size for OA as supply, with OA as demand with 1 values used as both supply and demand. There was a total of 100,721,296 pairs with 37,493,828 in distance pairs, and the test took an hour and three minutes to complete. This data size is not out of the question if one is investigating a large area and the tool should be able to cope with these data sets better.

The following attribute table (figure 3-16) displays results from the tests and it shows that without the ability to map the data it would be very difficult to read and interpret. This is why having QGIS was a benefit as it provided this functionality in an easy-to-use way. The results were as expected and the smaller catchment sizes got fewer results but typically had higher scores. As the catchment sizes got bigger, there were more results but the typical values decreased. This would continue until all points were within the catchment area, and then the score would be at the median value, which in this case is the number of LSOA divided by the number of OA, 0.190255.

	gid	code	Demand	Supply	500	2500	5000	10000	20000	30000	40000	50000
1	197	W00000068	1	1	1.151598	0.2059882870	0.2217268990	0.2128873250	0.2045178570	0.2206914000	0.1997990740	0.1944241970
2	1938	W00010037	1	1	1.006974	0.1621067520	0.1826876690	0.2032966790	0.2040330030	0.1969495400	0.1895733410	0.1880966290
3	27	W00000004	1	1	1.000000	0.6216428480	0.3363880320	0.2135554610	0.1742510600	0.1946836470	0.1781581410	0.1695392090
4	128	W00000040	1	1	1.000000	0.3858423220	0.2627561670	0.2134308310	0.2014098200	0.2164595690	0.1872024330	0.1804016780
5	208	W00000112	1	1	1.000000	0.6270354090	0.3189634380	0.1426208650	0.2049654950	0.1726205480	0.1487396140	0.1347846240
6	231	W00000314	1	1	1.000000	0.2365101780	0.2360181960	0.1805406600	0.1943139030	0.1988039730	0.2145848930	0.2125133770
7	300	W00000585	1	1	1.000000	0.2570362340	0.2102205010	0.1991788690	0.1701607260	0.1477685570	0.1519985210	0.1629933310
8	403	W00000435	1	1	1.000000	0.2672858500	0.2101092070	0.1390700500	0.1446576550	0.1874593980	0.1920505930	0.1909360580
9	504	W00000410	1	1	1.000000	0.3296028430	0.2590469220	0.1994874570	0.1808934640	0.1494157360	0.1235692720	0.1114121470
10	547	W00000592	1	1	1.000000	0.4678665210	0.3620659290	0.1888318070	0.1641721500	0.1611942930	0.1406075740	0.1419986300
11	657	W00001013	1	1	1.000000	0.7568699760	0.4476955190	0.2495378410	0.1470656960	0.1053429450	0.1185556500	0.1695319640
12	1068	W00001333	1	1	1.000000	0.5795931330	0.2035441190	0.1291785240	0.1808203650	0.1806028320	0.1994800630	0.2051693990
13	1100	W00001119	1	1	1.000000	0.2346644040	0.2084137710	0.1390464890	0.1439622020	0.1651580320	0.1791608780	0.1864959930
14	1173	W00001028	1	1	1.000000	1.0000000000	0.4601495970	0.2143081170	0.1375539340	0.1358396310	0.1543960390	0.1797983540
15	1380	W00001783	1	1	1.000000	0.1960376710	0.1146989280	0.1533144480	0.1637044890	0.1717198710	0.1834211870	0.1837746140
16	1451	W00001813	1	1	1.000000	0.2745186530	0.1231789950	0.1310734330	0.1605749080	0.1644578260	0.1762361230	0.1870508660
17	1490	W00001459	1	1	1.000000	0.3256954490	0.1527206650	0.1360791810	0.1764012540	0.1897546630	0.2052328110	0.2102577950
18	1743	W00001670	1	1	1.000000	0.2055190800	0.1477702050	0.1704039040	0.2313026730	0.2216356480	0.2110181260	0.2070547810
19	2219	W00002221	1	1	1.000000	0.2344299700	0.1368506730	0.1435281660	0.1908873330	0.1963119010	0.1878762660	0.1871755510
20	2441	W00002715	1	1	1.000000	0.2763767050	0.2174227060	0.2065081510	0.1939706610	0.1713025620	0.1495196480	0.1341700600
21	2479	W00002782	1	1	1.000000	0.2394052810	0.1563387670	0.1349398780	0.1750100780	0.1742848220	0.1727403240	0.1584124240
22	2516	W00002853	1	1	1.000000	0.2777363050	0.1278013730	0.1747581300	0.1947165640	0.1913839620	0.1794372430	0.1657141840
23	2706	W00002957	1	1	1.000000	0.3845431640	0.1494627690	0.1835744470	0.1710642930	0.1593252400	0.1688961340	0.1758073280
24	2715	W00003315	1	1	1.000000	0.5716065940	0.1905653910	0.2196728850	0.1837345600	0.1410757670	0.1461785430	0.1464004320
25	2748	W00003097	1	1	1.000000	0.3850489740	0.2977565940	0.1512265560	0.1862932140	0.1576008460	0.1635120470	0.1581433800
26	2793	W00003115	1	1	1.000000	0.5582038850	0.4053588700	0.1700461000	0.1493730350	0.1781037350	0.1928117190	0.1809841070

Figure 3-16 Copy of the QGIS attribute table

3.11 Key Findings from The Pilot Study

This proof of concept delivered a tool capable of implementing straight-line FCA analysis in an open-source environment. The result was a standalone program capable of performing accessibility analysis. Initial findings were shared with Tenovus in order to receive feedback on the design and functionality in relation to user requirements. In the development of the software, it was helpful to understand how the algorithm worked, and to highlight different areas for improvement in later iterations such as, performance and complexity. The way in which this program scaled shows that it was only suitable with limited datasets and the results using Euclidean distance were inherently less valuable than those which took into account the distance travelled. It also showed the need for more complex algorithms which took into account distance decay functions to give a more intuitive result. By understanding the way, the tool was coded it has been possible to narrow down future development routes and this assisted in focusing the search discussed in section 3.12.

3.12 Proof of concept: Limitations

However, using open-source software could be seen as problematic in that it left the tool vulnerable to change and could make the tool obsolete, which is, of course a problem with any software, but as there are rarely service level agreements in place it can make future planning difficult. This project used well-supported and long-term open-source tools to mitigate this risk. QGIS has been in use since 2002 and has improved year on year with additional functionality and user-friendly interfaces. It is used in organisations as a free and reliable alternative to ArcGIS that is too expensive for many users to consider. Python has been around since the early 1990s and is a very popular language, it is very unlikely to cease development due to the people and organisations that use it. The speed of the prototype tool was a major problem as it could not cope with large datasets in that format, and as any spatial database features were incomplete or difficult to access the solution was not a viable pathway. This needed to be addressed to make sure that the program was scalable and usable for the client. If this was not addressed it may not be used as often as possible, and any potential benefits may not be accessed by the client.

This tool used straight line distance to measure accessibility. This was not a good measure as we very rarely travel in a straight line to and from locations. A networked approach provided much more useful information which added weights to roads for speed limits and gave a much richer picture of the accessibility of different locations; but this was difficult to achieve in QGIS. A wider range of distance decay functions would be beneficial so that the user is able to pick the one which best suits their data, the decay functions were optimised for different types of data and this should be included in any documentation created to guide the non-expert users in their choices.

Data storage was not good enough, everything was completed on a local machine which meant that data needed to be passed from user to user and the reliance on shapefiles to supply the source data was not as efficient as it needed to be. This showed that a database could provide a very sensible alternative to help with the sharing and protection of the data, to make sure that original data sets are not overwritten or corrupted. QGIS proved to be a valid display mechanism and if PostGIS was used as it integrated well to provide full database and GIS functionality.

The tool was only able to compute the original 2SFCA scores and was unable to cater to any of the other methods that have been discussed. The other methods have their positives for certain size catchments and it would be good if this tool was able to provide more than just the original 2SFCA calculations.

3.13 Choice of Approaches

Many different alternative approaches could have been taken when building a tool of this nature. The use of QGIS has been justified in the previous section, but it is important to understand the other options that were available, and the reasons for them not being used at this stage. Google Maps API could have performed some of the required tasks; it was possible to write a script which could connect to their server and use the Google Maps routing capabilities to report far more accurate catchment sizes. This would have been a lot more difficult to implement initially and it was difficult to load the data in and get the results out. The way in which this would have to be constructed would have been a standalone program which connected to their API and then reported the results back to the program. Some of the main advantages of this are that Google would keep the network and the servers up to date which could negate a lot of future work for any potential user, and ensure the accuracy is as good as it can be. The negatives are that Google only provides this service free of charge for up to 2,500 calculations a day, and this would not work with the type of data that this project is using. As shown in the later study there are over 50,000 calculations even when the data used is on a small scale.

There is an almost endless multitude of pathways and combinations of software, support libraries, programming languages, development environments and databases that could have been combined to provide a solution pathway. What the prototype exercise showed is that QGIS needs easily accessible spatial indexing and built in routing solutions; any language could be used but python, java, C++, C# are the strongest contenders. The use of QGIS was justified in the previous section but it is important to understand the other options that were available, and the reasons for them not being used at this stage

For example, the Google Maps API could perform the networking tasks that were required; it was possible to write a script which would connect to their servers and use Google Maps routing capabilities to report far more accurate distances or times. The way in which this would have to be constructed would have been a standalone program which connected to their API and then reported the results of network calculations back to the program for further computation to yield the 2SFCA results. Some of the main advantages of making use of the Google Map routing API were that Google would keep the network datasets up to date, which could negate a lot of future work for any potential user and ensure that a high level of accuracy is maintained as well as being able to report public transport routes and times which could have potential future benefit. There was no spatial indexing to rescue this computational load which would be inefficient in a project such as this, as the datasets can be very large and the ability to use spatial indexing is critical to computational performance.

Finally, using a third party such as this moves away from the free and open-source philosophy of the project and it would mean that if Google was to exit the market, or change terms and conditions, the project could be rendered useless. There are options other than Google for using a free API like OpenStreetMap and Garmin, with additional concerns regarding the quality of the networks and maps which may vary between provider.

Another alternative would have been to use ArcGIS, which was the largest proprietary GIS on the market, or one of the other proprietary systems that could be used to create this tool. This would have been easier than using QGIS due to the increased functionality of the software and the ability to create networks within the software itself. The problem with this solution was that it required users to purchase the software, which can typically prove to be very expensive. Again, it went against the FOSS philosophy that the project was hoping to adhere too.

Given these concerns and limitations the preferred choice would be to use a spatially enabled database such as PostgreSQL/GIS, which is a free and open-source database used by multiple large organisations and government authorities worldwide, with a routable network to compute the calculations. This can still be used with QGIS to provide all of the additional functionality that QGIS has to offer. By making use of PostGIS spatial indexing it was also possible to use the r-tree function to quickly and efficiently remove the vast majority of the calculations that were being completed in the pilot program by removing calculations that were clearly outside of the parameters set, the r-tree function discarded many of the data points that were likely to be outside of the parameters and were therefore not processed, easing the burden on the hardware to compute as many calculations. There are other benefits to adopting this pathway in the future, including the data protection and performance enhancements associated with using a fully-fledged database. PostGIS includes a library called pgRouting, that provides an easy-to-use alternative to creating and using routable networks from line/node datasets.

It would have been possible to build a complete standalone piece of software able to complete all of the desired tasks from scratch. That would however have been considerably more difficult and time consuming to implement than making use of libraries, and data management systems such as PostGIS/pgRouting, and the options for how this could be implemented are almost endless. Using C++, QT and openGL would be one example. Pursuing this pathway would have made sense if there was a specific task that a private company wanted to achieve, and they wanted to ensure that no-one else was able to view or benefit from their findings, as it would be completely private and built for that one purpose only. It is a possibility for future work but there needs to be further investigation into possible alternatives to make sure that work is not being replicated unnecessarily.

Such possibilities are just a few of the alternatives that have been investigated so far and there are more products and tools released periodically which are tested and evaluated to see if they can help in creating the final iteration. It is important to remember the philosophy of the project and the user requirements to ensure the project stays focused on creating a usable and well-designed piece of software which can be used by non-experts and benefit organisations such as Tenovus in planning the provision of their services.

3.14 Chapter Summary

The prototype software solution described in this chapter showed that it was possible to create a free and open-source tool with the ability to provide straight-line 2SFCA scores. Building this as a plugin to the QGIS system had a number of benefits, in particular allowing the results to be manipulated and reported in a visualised format in a simple way. The reliance on Python and PyQT provided a productive environment for the code developer, and enabled the complex calculations needed for a 2SFCA analysis to be performed in less than 500 lines of code.

A user-testing suite needs to be designed to ensure that the end user is the focus of the design. It is important for the free and open-source philosophy to remain central to all the ideas and tools that are being considered, and only those that fit this philosophy will be included.

Data is very important to the project, and for this tool to provide meaningful results it will ultimately need to use much more detailed data than that tested in the pilot phase, in order to provide insightful results to the user. This implies that the performance of the tool needs to be improved to ensure that it is scalable and will be able to handle any of the data sets that it may have to process in the future in a full production environment. It has become clear through the pilot study that it will be necessary to integrate a spatial database and use its spatial indexing capabilities to minimise the computational load associated with the distance calculations that form the heart of the 2SFCa algorithm.

The maps produced by this pilot study were able to show that a tool of this type can be of benefit to workers at Tenovus who would be able to visualise the impact of alternative strategies and delivery site choices on the communities that they serve. The adoption of the 2SFCA accessibility score was a good choice for this project as it had been demonstrated to be of value in a range of healthcare and cancer specific studies and provided detailed supply to spatial demand ratios that can be used by Tenovus when they require further funding or to access the implications of using alternative mobile site locations. The linear distance decay function was also a good addition as it helped to provide more accurate and meaningful scores, but also a wider range of available distance decay functions should ideally be included so that the most applicable function can always be used for the analysis that is being undertaken.

The prototype provided a number of useful insights into design decisions, potential pathways, absolute needs and optional niceties, that can be used to guide the on-going main phase of the project. Specifically, we learnt that a different approach is needed, it has been shown that the program does not scale well and that the use of a spatial database is important for performance as well as data storage and manipulation. The program clearly needs to easily access routing capabilities as well as to deliver more complex algorithms which can offer a more detailed understanding of accessibility.

Chapter 4 Methods

4.1 Introduction

The aim of this chapter is to explore the methods used to create the final iteration of the tool, and demonstrate the ways in which this has been tested and applied to a range of data sets. This includes a discussion of the development environment and the software and hardware used in this study, and incorporates an in-depth investigation of potential FOSS solutions and a full justification for the choices made. The potential for a range of routing algorithms is discussed prior to a fuller review of the different FCA algorithms which can be implemented within this solution. Several case studies were been completed to show potential users how the tool could work and benefit them and there is a discussion regarding display options. Finally, the documentation that was developed alongside the implementation is reviewed and discussed.

This chapter looks in detail at the software and hardware options available to develop the tool. All of the software used was FOSS and can be accessed by anyone wishing to replicate this tool or create a similar one. The hardware used in this project is good quality but familiar, and consists of a non-specialist laptop and desktop which should be widely available. Microsoft Visual Studio has been used as an integrated development environment (IDE) utilising C#.Net for the user interface (UI) with PostgreSQL/GIS as a spatial database and the pgRouting extension for network analysis functions.

There are various options available when creating this type of application so it was important to make informed decisions regarding these. It was necessary to evaluate not only the quality of the tool but also its ease of use and the developmental knowledge required to implement it; as the tool will be passed to a team of non-experts to use and may require updating in the future. This section outlines several possible solutions and a justification of the final choice of software tools used to implement the FCA models. There are several ways to set up the routing analysis depending on the systems available and the desired output. Similarly, there are a large number of potential algorithms that have been developed in the FCA literature (reviewed in Chapter Two). Such research has sought to improve the algorithm and make it more appropriate for analysing a wide range of datasets. The outcome is that researchers have considerable choice when deciding which algorithms to implement within the tool to make it as accurate as possible whilst trying to ensure usability for non-GIS experts.

After discussions with potential users of the system within Tenovus, in order to demonstrate the usefulness of an FCA approach to analysing spatial patterns in access to cancer services, it was decided to develop several case studies that used real data to demonstrate the implications of moving one of their mobile units or adding additional resources to a new area. To do this supermarket data that incorporated a measure of parking space size was utilised and several maps were created to show the difference in accessibility under a range of different scenarios. To visualise the results QGIS was used but this was mainly due to the preference of the author; the tool created 2 files that can be used in QGIS or displayed in other available GIS packages. This tool is very simple for the end-user and requires a small degree of user-input in order to implement some quite complex calculations. It is therefore important that these processes are documented to ensure the tools are successfully implemented. Therefore, three different documents have been created (appendix C):

- A User Guide which shows the end-user how to input data that can be used to implement these models, and which demonstrates how to visualise the results through a range of outputs.

- A superuser manual which shows the more proficient user how to download and update all of the software used, and also how to add data to the database and all of the more complex tasks that are required.

- A guide to FCA which explains how the calculations work and what the results mean, to assist in interpretation.

The quality of the documentation determines how often the tool is used and whether it is a success and has been evaluated in the usability studies documented in Chapter 5.

4.2 FOSS considerations

There were a number of key considerations associated with this project; firstly, for reasons outlined previously, an open-source solution was required to implement these models so that they could be freely used and customised by those that need to access it. To do this involved using and evaluating a number of different software solutions and implementing those that were the most appropriate for the project. The following section summarises the respective advantages and limitations of some of the options available.

To complete the 2SFCA calculations there were several key tasks that were required of the software. A key requirement was the ability to sort through voluminous spatial data in an efficient way on a standard computer. A routing tool that is flexible, which can traverse multiple routes efficiently, provides more accurate 2SFCA results, and is key in delivering an accurate tool as well as providing important results from the data gathered in this project. An interface that is easy to understand by non-experts was vital to delivering a usable tool. The software needed to be flexible and provide a stable development platform which would be usable for extended periods of time to ensure the tool does not have to be re-designed at every major update.

4.2.1 Development options

Table 4.1 compares several different solutions that were considered for this project. The list is not exhaustive but comparing these alternatives provides a good overview of the current capabilities of each of the options as well as their strengths and weaknesses. One of the aims of this project is to understand the different ways in which the task can be completed and to find the ‘best’ solution. This solution is subjective and as can be seen below, there was a wide variety of ways in which these tools could have been implemented. By drawing on a description of the functionality of a range of different FOSS options, the aim here was to compare and contrast the capabilities of such software tools and decide the most appropriate solution for this study in the light of the project aims identified in Chapter One.

	QGIS	**Stand-alone program (e.g., C++, QT and openGL)**	**GRASS**	**PostGRESQL/GIS and pgRouting**
FOSS	Yes	Yes	Yes	Yes
Development complexity	Using Python and QT creator it is a suitable environment to develop in. This	To create a stand-alone program is a very complex task and would involve developing tools	It is possible to develop add-ons for GRASS by using Python or a number of other scripting	Using SQL and utilising the built-in capabilities of PostGIS and pgRouting provided

	solution is quick to develop and the stacks cater for quick UI development.	and functions that are already available in other ways such as linked libraries and passing information between several component software applications.	languages, and it was felt this could present a workable development environment.	a suitable development environment. Many options are available for UI development as there is not a stack set up for this.
Network analyst capability	There is direct access to a network analyst within QGIS. It is possible to use the networking capabilities of the GRASS plugin however. The network analyst is not used often and there is limited documentation regarding its usage.	These functions would need to be created or imported and amended from suitable libraries (i.e., RoutingKit). To create a flexible routing program from scratch would take a large amount of development time and is unnecessary as solutions already exist that can be called upon.	There is access to network analyst functions from within GRASS.	pgRouting provides many advanced algorithms which can be accessed via SQL scripts issued from a front-end program.
Speed of solution	The QGIS solution could work but would need to use a spatial database to have acceptable performance.	This solution has the potential to be very fast as it would be created and optimised for this sole purpose.	This solution could be suitably fast as it uses a spatial database.	This has the potential to work within a tolerable timeframe due to the access to spatial indexing and bounding boxes.
Future	This is a sound	Would require a	There are many	This is a sound

development	platform used by many developers and has the potential for further development.	good knowledge of the tools and libraries used in development to develop further.	add-ons already created and any tool created has the potential to be developed further.	platform used by many developers and has the potential for further development.
Stability	QGIS is stable and has been in use for many years.	As a stand-alone piece of software, it would be stable as it would not rely on other software to work.	This is one of the oldest GIS and is a stable environment for development.	The software used in this solution is well established and managed. It is used by some very large organisations and should provide a stable environment.
Visualisation	It is easy to visualise the results with QGIS. It has the ability to then customise the map and save it in different formats.	There is the potential for visualisation. Although the results may need to be transferred to a GIS to customise and create good quality maps.	There is the ability to visualise the data within GRASS. It is not as user friendly as the QGIS option.	There is no built-in way to visualise the results. QGIS or similar would need to be used to visualise the results.
Benefits	An easy development environment and the ability to visualise the results in one piece of software. It also offers all of the other benefits that a GIS has to offer. Provides a good environment	A specific and precise solution for E2SFCA which has the ability to be tailored to other projects. It could be quick and stable.	A robust and well-established FOSS GIS which has many options to develop add-ons. The ability to visualise the results in the same program helps to reduce movement across platforms. Having built in	The use of PostGIS enables large data sets to be stored and managed centrally as well as manipulated in a quick and efficient way. The development environment is easy to use. pgRouting offers many different

	for the non-expert user.		spatial database and network analyst functions is a real benefit.	routing algorithms and both platforms are stable.
Problems	Although QGIS could provide a solution it would need to use a spatial database to have acceptable performance. The networking capabilities are not as complete as some other options but this could be overcome.	The development complexity for the tool would be high and there are many solutions available which could be used instead.	This solution has the potential to work well. Using GRASS may not be as easy for the non-expert user as some other solutions available.	This solution has the potential to work well. Not having a GIS to display the results is not ideal, but any GIS could be used or the results could be displayed online using a variety of FOSS solutions such as MapServer, GeoServer or OpenLayers.

Table 4.1 Comparison of available development options

4.2.1.1 QGIS

QGIS is a is a complete desktop GIS solution with facilities to store manage, capture and analyse various spatial data types. QGIS started development in the early 2000s and was released in 2009. QGIS is available on most platforms and can be used in conjunction with other FOSS software (PostGIS, GRASS). It is used in organisations (local councils, US National Security Agency) and provides a number of different functionalities (Raster & Vector layers). It has been used in academic publications (Ranga and Panda, 2014; Kim, Byon and Yeo, 2018) but not as widely as ArcGIS. QGIS is under the OS GEO banner and can be downloaded with optional extras that can tailor the capabilities for whichever purpose.

An initial tool was created using QGIS, QT creator, QGIS plugin builder and Python (Methods 1). There were advantages when using QGIS; mainly, it was easy to use and had the option of visualising the data all in one system. QGIS was a stable environment and had many of the advantages that ArcGIS had as it is a full FOSS GIS. The challenge of working in this environment only was that it did not currently have a capable network analyst, and an external solution (GRASS, PostGIS) needed to be used. QGIS proved to be inefficient in how it was able to perform the calculations and this affected the speed of the program when dealing with medium or large data sets. By using a spatial database to manage and store the data it should be possible to increase performance, but this solution was less efficient than using the database to perform all of the tasks, and then visualizing the results using a GIS.

4.2.1.2 Standalone system

There were a range of ways that a stand-alone system could have be created for a tool such as this. One of the most popular languages is C++ but Python or Java could have been used just as easily. The main considerations with this approach were the libraries that were available to help build the functionality. Creating a tool from scratch using C++, QT creator and openGL was a viable option for this task. Creating a tool in this way enabled the most efficient algorithms to be used and a tool could be created for this specific task. The problem with this solution was complexity, it was a more difficult programming environment and it would have been difficult for the program to be modified for alternative tasks. This would have also involved repeating approaches adopted in existing programs.

4.2.1.3 GRASS

GRASS GIS is a full FOSS GIS and has a similar level of functionality to QGIS in that it is a full desktop GIS. It has come from a background more focused on raster data and is slightly less complete in its vector data capabilities. GRASS began development in the early 1980s and has been continuously developed since then. It has been used by a number of US agencies as well as private companies and universities and has been well represented in academic journals (Leal et al., 2003; Scavuzzo, Provensal and Lanfri, 2017). GRASS has various capabilities and can analyse and display raster and vector data. It also has a network analyst function and works well alongside QGIS.

GRASS offers a credible solution for creating the tool. It has a network analyst capability and there are a number of development options available for creating an add-on. The ability to visualise the results in the same environment is helpful and would limit the use of multiple platforms. GRASS therefore represented a potential option for implementing this solution. The user interface was not as user friendly as QGIS and as this tool is to be used by non-experts, a more simplistic environment would be advantageous.

R would also be a suitable candidate for a project such as this. The development of the spatial analysis libraries and the ability to visualise results in one place make it a strong option for problems such as this. At the time R this research was conducted was much less prevalent and was not as complete in terms of the libraries within but the progress made within R would mean that it should be considered for future projects.

In summary, there was a high degree of functionality available in FOSS solutions and the number of platforms was encouraging for an open-source solution. The respective advantages and limitations of some possible contenders were considered and evaluated above for their potential application within the rest of this study. A full justification for the chosen solution is now outlined in more detail.

4.3 Final choice of development environment and solution stack

This final tool was developed using 2 computers, a laptop (i7-6500U CPU @2.50GHz, 16gb RAM and 500gb SSD) and a desktop (i5, 16gb RAM and 1tb HDD); both running Microsoft Windows 10. The user interface (UI) was developed using C#.Net with Microsoft Visual Studio 2017 as the IDE. The tool uses QGIS (v2.14.14 – Essen), PostgreSQL (v9.5), pgRouting (v2.3.2) and PostGIS (v2.3.2).

The updates that have occurred since this stack was assembled are themselves backwards compatible so the tool should work with more recent versions of any of this software. All of these components are FOSS and are readily available for download. Using the OSGEO4W tool it is possible to download all of the components from a single location at the same time and the user can select the version of each that is required. The reason for adopting version 2 of QGIS is that this has been the most stable and most used version at the time of development with version 3 being more commonly used from mid-2019. The data produced from the tool will work with the newer versions of QGIS and could be imported into most GIS.

Microsoft Visual Studio and C# have been selected as they best integrate with the experience within Tenovus. This choice was taken as Visual Studio is now compatible with most platforms and is a very popular solution which although not open source is free to download and use. The benefits of Tenovus being able to make changes in house made this selection appropriate and all of the power of the program is within the SQL statements which can be easily implemented in a different language for UI creation if someone wished to work in another environment.

4.3.1 Post GIS/pgRouting

PostGRESQL is a FOSS relational database management system which has been in development since the 1980s. It can be used across platforms and provides solutions for large and small projects. PostGIS is an extension for this database and allows spatial databases to be used. The use of geographical information in a database provides a very powerful tool to compute large datasets in an efficient way. PostGIS is widely used in by other organisations such as ArcGIS, QGIS and GRASS. It has been used in academic research (Bhaumik et al., 2012; Singh, Singh and Singh, 2010) across multiple specialisms.

pgRouting is an extension for PostGRESQL/GIS and allows for different routing algorithms to provide routes from the data in the database. This can then be displayed using QGIS or another GIS and the information can be used to gain further insights into the distances travelled rather than a standard straight-line distance. PostGRESQL/GIS and pgRouting provide a large amount of functionality; the ability to store and manage large datasets provides a good foundation for this tool. The use of spatial indexing and bounding boxes allows for good performance and the different routing algorithms within pgRouting give this option flexibility. There is no ability to visualise the results without using an external GIS, but it is simple to produce an output which is ready for visualisation.

As well as the options listed above there are several others that were investigated during the course of this study such as gvSIG, Open-Source Routing Machine (OSRM) and the OpenRouteService (ORS). These tools provided different alternatives to those above. gvSIG is a FOSS GIS similar to QGIS and has been predominantly developed in Spain, it has functionality similar to GRASS and would provide a good option for the tool. OSRM and ORS are services which can be used to provide routes and as they are FOSS they can be hosted externally and provide routing information. If the tool were to be used on a larger scale a solution like this could be advantageous as are the availabilities of different modes of transport.

In summary, there is a large amount of functionality available in FOSS solutions, and the number of platforms is encouraging for an open-source solution. The respective advantages and limitations have been considered and evaluated for their potential application within the current study and in the rest of this chapter, a fuller justification for the chosen solution is outlined in more detail.

4.3.2 Algorithm choices

There have been a number of developments since the initial FCA formulation (Wang, 2000) including those proposed by Dai, (2011) and Luo (2014). These were described in more detail in Chapter 2. These advancements, whilst claiming to provide a more realistic measure of accessibility, have also made the calculations more complex and in some cases additional parameters are required. It would have been possible to fine tune a calculation using all of the latest methods to work well with the data that has been provided from Tenovus, but this may make the tool much more difficult to use, and as it is designed to be used by the non-expert usability is very important. Three different algorithms have been implemented into the tool to reflect this; the original 2SFCA method is implemented (figure 20) which is similar to that of Wang, (2000) together with a linear distance decay function (figure 21) similar to Luo, (2014) and a Gaussian distance decay function (figure 22) similar to Dai (2011).

4.3.2.1 No Distance Decay 2SFCA

As discussed in the literature review one of the main issues with the original 2SFCA method was that it assumed everyone within the catchment area had the same access (McGrail & Humphries, 2009). This means that if there is a 30km catchment the calculation assumes that those who have to travel 1km have the same access as those who are 29km away from the demand point. This method is still useful and shows general patterns but it is possible to increase the accuracy of the algorithm which can have significant effects to the results. Figure 4-1 shows the results of a 2SFCA calculation.

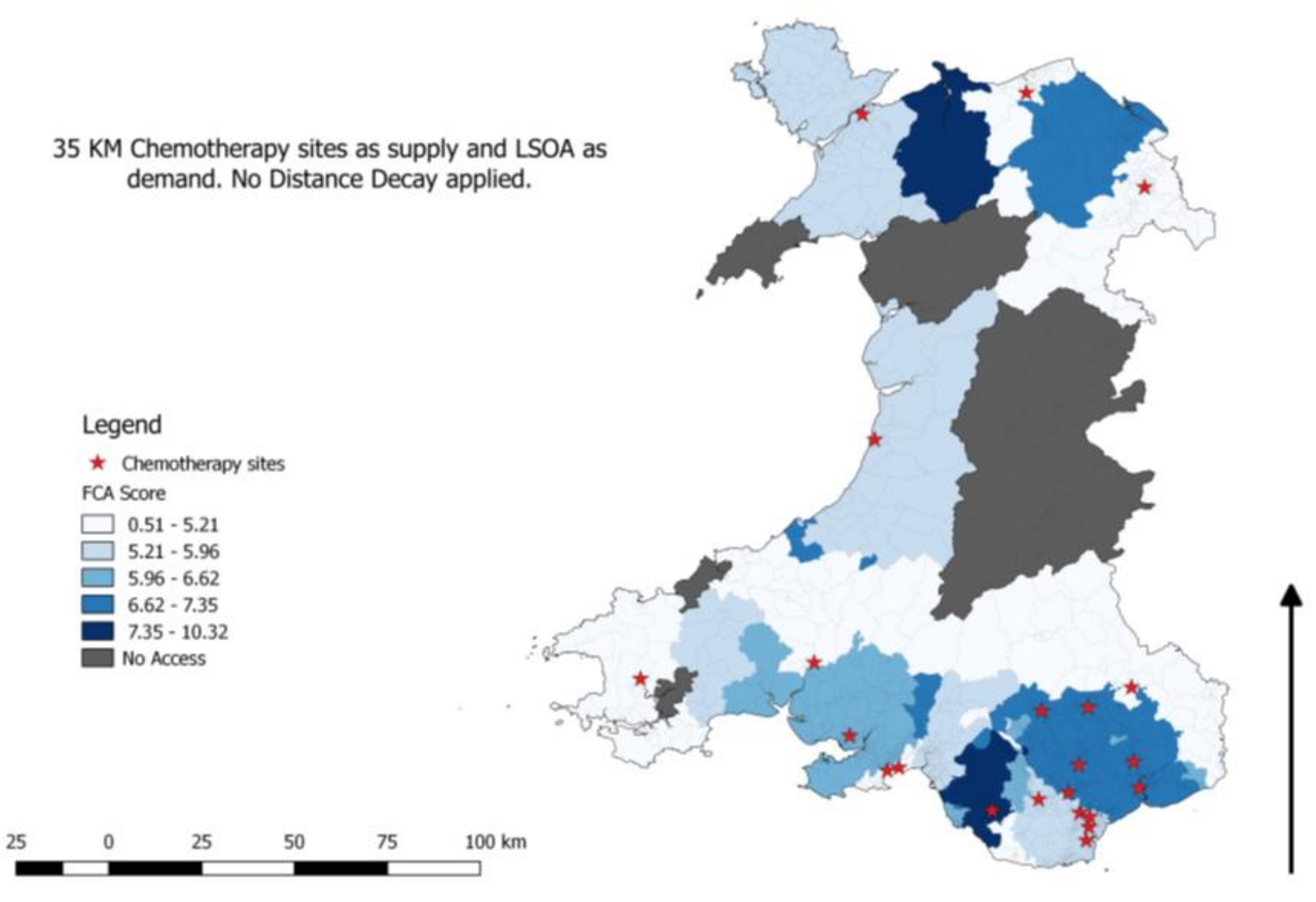


Figure 4-1 2SFCA results of Chemotherapy sites as supply and LSOA centroids as demand with no distance decay

4.3.2.2 E2SFCA (linear)

The use of a simple distance decay function had the ability to scale the access of those within the catchment area; the linear distance decay function alters the scores on a sliding scale from 1 at the datapoint to 0 at the edge of the catchment area with 0.5 for a point half way between as shown in figure 4-2. This allowed for a much more detailed understanding of the accessibility within catchment areas and was useful in calculations with larger catchment areas (Luo and Qi, 2009). Figure 4-3 presents the results of the E2SFCA with a linear distance decay.

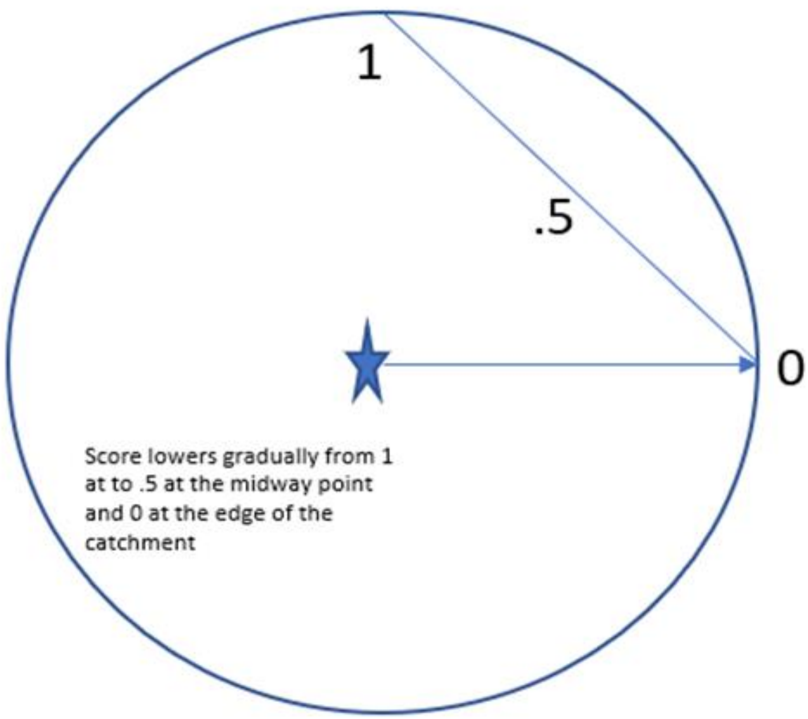


Figure 4-2 Diagram showing the linear distance decay function

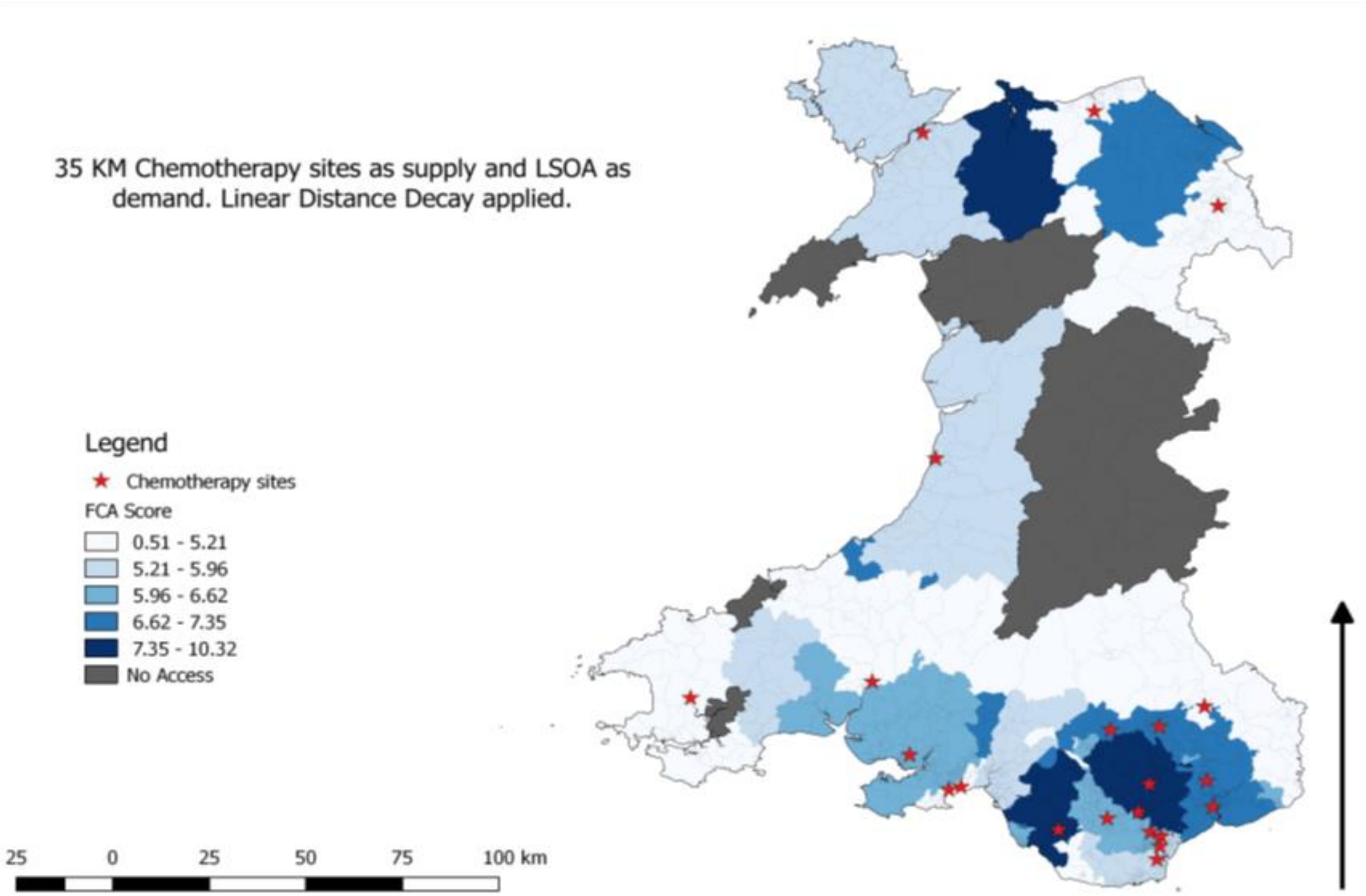


Figure 4-3 E2SFCA with a linear distance decay results of chemotherapy sites as supply and LSOA centroids as demand

4.3.2.3 E2SFCA (Gaussian)

After reviewing the user data gathered from Tenovus (section 4.3.2) it was possible to see that the data was Gaussian in nature, and this led to the choice to implement the Gaussian decay function too (see Polo et al., 2013). The use of a Gaussian distance decay suggested that those living within a set tolerance had similar access and this then gradually tailed off to a plateau as shown in figure 4-4. Figure 21 shows the results from the Tenovus users, and this is explained in greater detail in section 6.5. By utilising the data below, it was possible to set the catchment area at 35 km and give those within 8 kms a higher accessibility score than those at larger distances. This could be different for users other than Tenovus and can be adjusted to suit their needs. Figure 4-5 shows the results of the E2SFCA with a Gaussian distance decay.

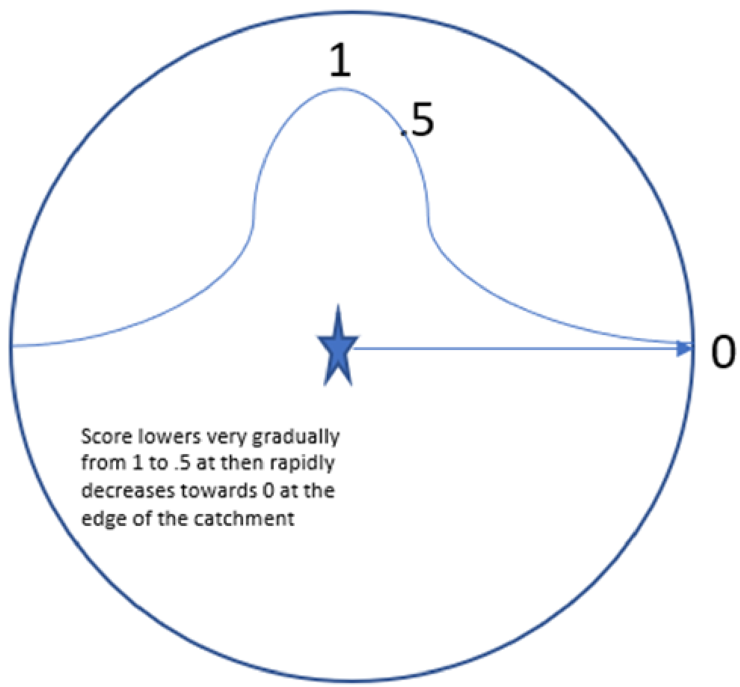


Figure 4-4 Diagram showing the Gaussian distance decay function

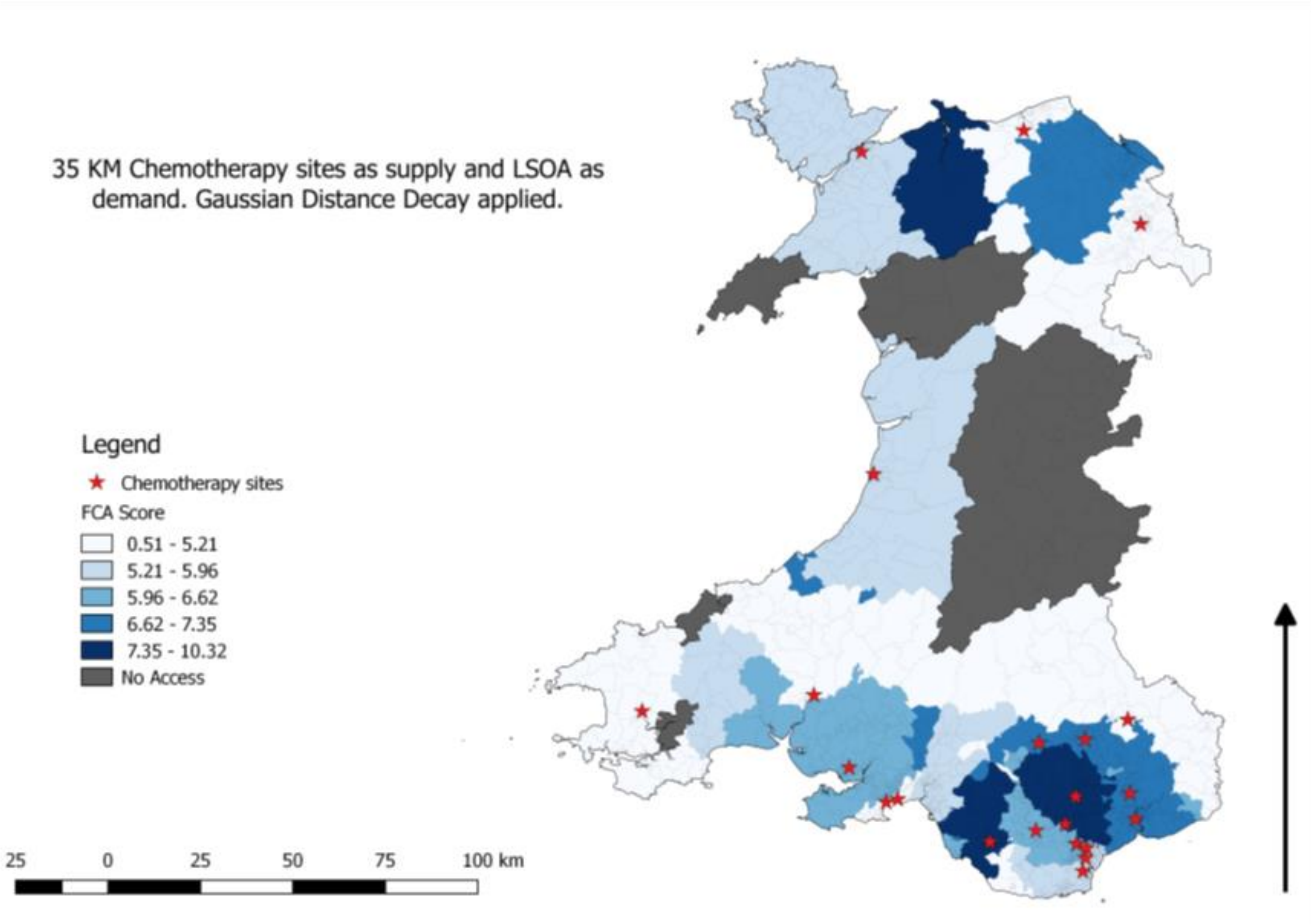


Figure 4-5 E2SFCA with a Gaussian distance decay results of chemotherapy sites as supply and LSOA centroids as demand

4.3.2.4 Connecting to the datapoint or to the network

Understanding the effects of small changes in the design of the tool was key to this project's success. One way in which the project is able to contribute to science is to test the differences in how these programs are built, and the effect that these have on the results of analysing real data. The approach taken was to build a tool which offered a number of algorithmic implementations. The solution developed is able to (i) connect to the nearest node on the network and (ii) connect to the nearest line on the network. It is less computationally demanding to connect to the node whilst it is more accurate to connect to the nearest point on a line. Understanding the degree to which the choice might affect the outputs produced is important, because if there is very little gain for the extra computational work it is questionable whether the extra computation is worthwhile. The data sets being used had a bearing upon this issue as the OSM dataset works better when connecting to the line than the OS Open Roads data set. This is due to the way the data is provided and the different functions used to connect to the line, or to the node within the program. When designing a tool for non-experts it was important to understand the effect that these decisions had and then implement the most suitable solution (see section 6.4.3 for results of testing).

Although connecting the datapoint to the nearest line was more accurate, it was important to understand that most of the datapoints are estimations and not based on an actual place. This type of collated data that is used in the LSOA population weighted centroids is not accurate, so would a small change in where the datapoint intersects with the network have any bearing on real world data and the results? It was important to present the most accurate results possible, but it was important to understand that even if the tool was accurate to nth degree the results would only be as accurate as the data which is being processed. A comparison has been completed and is described in more detail in section 6.4.3.

4.3.2.5 Routing choices

There are ways of traversing a graph to find the optimal route and within pgRouting there are 11 main routing algorithms which offer slight differences in how this is achieved. These include:

- The Dijkstra algorithm – uses edges of a graph to get the shortest path between two nodes and provides a detailed route.
- Turn restricted shortest path algorithm – based on Dijkstra it is able to take into account real world problems such as roads with no entries and one-way streets.
- Travelling Sales Person algorithm –finds the best route between multiple target nodes and the source node to show the most efficient way to travel between multiple nodes and return to the source.
- Driving distance algorithms – uses Dijkstra to show all target nodes within a specified distance.
- The Floyd-Warshall algorithm – normally used for dense graphs; returns the sum of all costs between each pair of nodes on a graph.
- The Johnson algorithm – normally used on sparse graphs; returns the sum of all costs between each pair of nodes on a graph.

- The A* algorithm – an efficient routing algorithm that analyses the graph prior to the search and can perform well if the network graph is small, when used on large networks the initial analysis can be inefficient.
- Bi-directional A* algorithm – traverses the graph from the both the target and the source node.

The Dijkstra algorithm was utilised in this study as it did what was needed and did not require any precomputation or heuristics to work. This tool could be feasibly used on a continent sized network, and the Dijkstra algorithm is the most suitable for this. The way in which the Dijkstra algorithm finds its shortest path is relatively simple; by specifying a source and target node the algorithm works to find the path of least cost between them. To do this the distance between intersecting nodes is recorded so that the path with the lowest total cost can be reported.

Within pgRouting there are several Dijkstra functions that could have been used:

- Dijkstra cost algorithm – uses the Dijkstra algorithm but only reports the total cost of the route.
- Bi-directional Dijkstra algorithm – traverses the graph from the both the target and the source node.
- Dijkstra cost matrix algorithm – uses the Dijkstra algorithm to produce a cost matrix with the total cost of the journey between all pairs.

The Dijkstra cost function was utilised for this project as the sum distance was the only one of interest, and although many of the other algorithms could work, they are not as efficient in this setting. The driving distance set of functions offered a solution, but it was difficult to separate the target nodes and the standard nodes, and provided a slower solution in practice.

4.3.3 The final iteration of the tool

This tool was developed using 2 computers, a laptop (i7-6500U CPU @2.50GHz, 16gb RAM and 500gb SSD) and a desktop (i5, 16gb RAM and 1tb HDD); they are both running Microsoft Windows 10. The user interface (UI) has been developed using the C#.Net framework with Microsoft Visual Studio 2017 version 4.7.03056 as the IDE. The tool was developed using QGIS V 2.14.14 – Essen, PostgreSQL V 9.5, pgRouting V 2.3.2 and PostGIS V 2.3.2. The updates that have been made are backwards compatible so the tool should work with more recent versions of the software. All of these tools are FOSS and are widely available for download.

Using OSGEO4W tool it is possible to download all the tools from the same location at the same time and users can select the version of each that is required. The reason for using version 2 of QGIS is that this has been the most stable and most used version at the time of development with version 3 being more commonly used from mid-2019 but the data would work with most GIS.

Figure 4-6 shows how the different parts of the stack interact. The tool works by utilising PostgreSQL/GIS to perform the calculations which are controlled by a UI developed in Visual Studio using the C#.Net framework and accesses the pgRouting extension in PostgreSQL for routing calculations. The program has around 17 SQL queries which order the data, perform the calculations and provide a table which is easy to visualise.

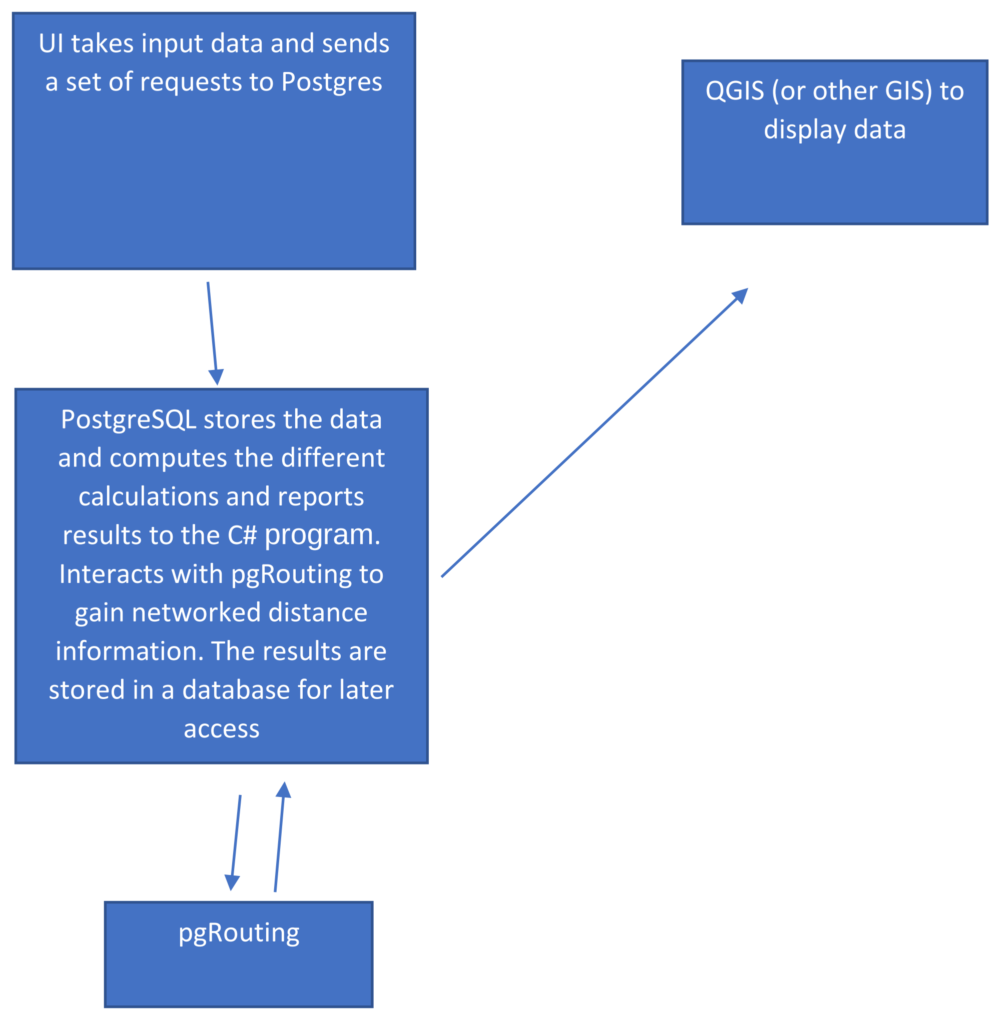


Figure 4-6 Stack overview

The tool works by utilising PostgreSQL/GIS to perform the calculations which are controlled by a UI developed in Visual Studio using the C#.Net framework. The program has around 17 SQL queries which order the data, perform the calculations and provide a table which is easy to visualise. Figure 4-7 walks through the SQL queries used to perform the 2SFCA calculation and show how the program works. There are several adapted versions of this code which feature distance decay functions and other ways which the data points are attached to the network but they are expanded on in later sections. The following flow chart shows the basic steps used to perform the calculations. C#.Net is used to select the correct data sets, choose the algorithm and input the catchment size. This program then runs the SQL commands which manipulate the data in the spatially enabled database.

Figure 4-7 2SFCA Algorithm workflow

Initially the tool needs to find a position on the network for each of the data points. This can be done in a couple of ways and is discussed in more detail in Section 6.4.3. Most data points do not sit on the network itself, so establishing the most effective way of assigning their position was an important task. Figure 4-8 shows that there were several choices that could have been made for connecting the point to the network. There are 3 nodes that are a very similar distance from the point and the closest place on the network requires a connection to the line connecting the nodes. The choices made here could have an effect on the overall performance of the program and the quality of the results. This method also had to assume that the access to a road network was made at the closest point which may or may not be the case depending on the building, and hospitals often have multiple entrances and exits. The algorithm works whether the network is connected at the nearest node or to the nearest place on the network.

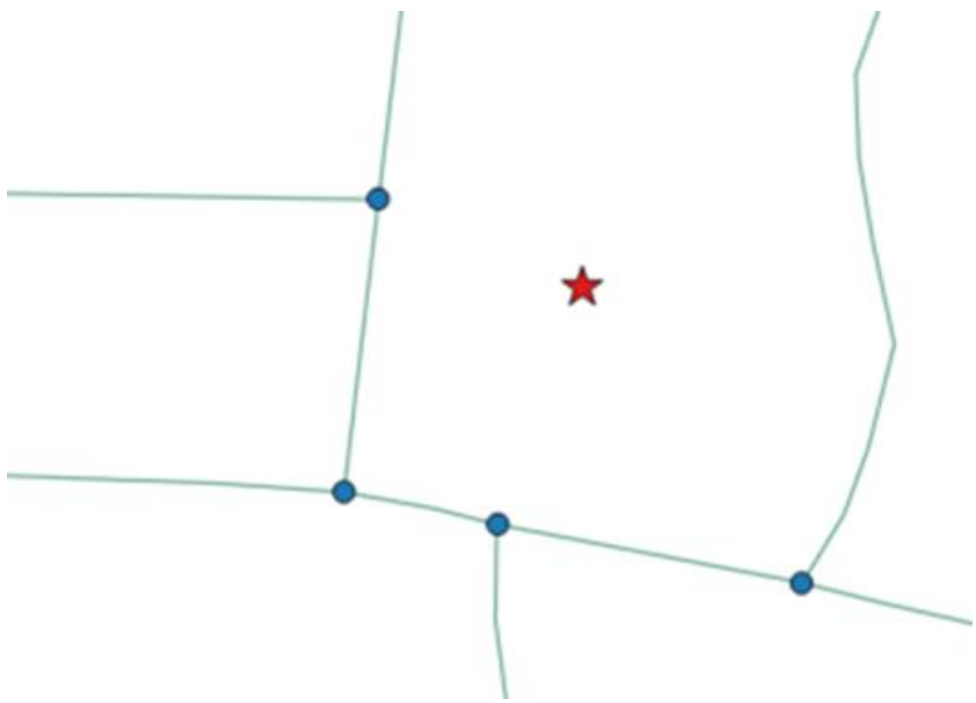


Figure 4-8 Centroid with possible connections to the network

One way in which this can be achieved is to find the nearest position on the network to the supply/demand point, then split that line and rebuild the network with the position noted as a new node. This approach would be particularly well suited to relatively static data sets that are unlikely to change, such as the data relevant to Tenovus. This approach would save on computation each time the program runs, but it adds to the complexity for the end user and splitting and rebuilding a network can be difficult to complete correctly. The simplest solution is to find the nearest node on the network, as this can be done using a function, but can be very inaccurate. Assigning the datapoint to the nearest position on the line is more complex and requires more computation to achieve, but it also provides the most accurate solution. It should be noted that the nearest position on a line may not be correct, but when working on data of this size it would be unrealistic to check every datapoint used. The tool has the option to both connect to the line and to the node, and it has been used to perform some preliminary investigations into potential implications of these options on resulting 2SFCA scores (see section 6.4.3).

Several tables are created by the program to organise the data for the calculations, the data needs to be presented in the correct format as there are multiple data sources; the tables only take the information that is required for the precise calculation. This could have been done using temporary tables or views. The choice to use tables instead of a view or a temporary table allowed for each table to be saved and inspected, there are very little advantages within PostgreSQL to using views or temp tables compared to some other databases, and this could be changed if a user would prefer at a later stage. Using a spatial database to sort the data in this way assisted in the performance of the tool. The first iteration of the tool had to loop through every datapoint, whereas this solution enables a large amount of data to be sorted much more efficiently. Once the nearest point on the node is stored the tool performs a Euclidean distance search around each datapoint and removes any outside of the selected distance. The ST_DWithin function (figure 4-9) is used to isolate nodes within a specified straight-line distance, this function is used over the ST_Distance function (unable to use with spatial indexing) and the ST_Buffer (less accurate). This allowed the program to limit the number of routes that it needed to calculate and improved the performance of the tool significantly.

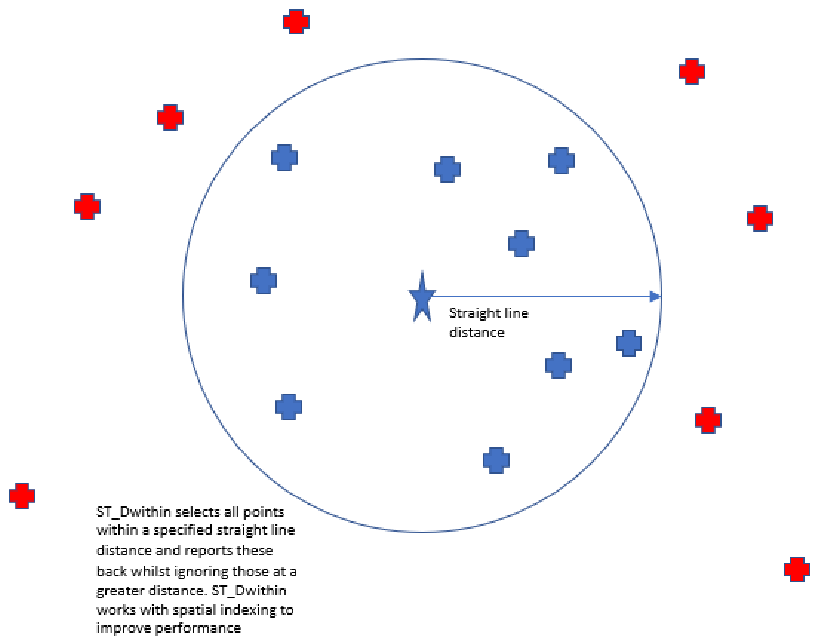


Figure 4-9 Example of ST_DWithin function

When performing a distance calculation using pgRouting, the way the data was presented determined how the program ran. It was possible to run one to one distance, one to many, many to one and many to many (see table 4.2), and this choice had a dramatic effect on performance as the one-to-one worked much like the proof-of-concept tool and returned one route at a time, this was not an efficient way of computing and was mainly used to find a route as opposed to finding many different routes. The one-to-many choice was more efficient as it took a single input as the starting position and then accepted an array of numbers and reported all of the routes found from the single entry to each one in the array, this utilised the database and computed dramatically more efficiently than the one-to-one method. The many to many method reported a route from every input and in the case of this program many of these routes were not required. To utilise the one-to-many functions the data needed to be configured appropriately, which is to have each supply node as an individual point (the 'one') and all nearby candidate demand nodes as an array (the 'many'). Using this option removed a lot of additional processing compared to the one-to-one API call, and consequently the program operated much more efficiently as it was doing considerably less calculations. The effect that this had in testing took the process from over 30 minutes or hours, to less than three minutes. This process had a significant time gain for medium to large datasets and allowed the program to operate at a more acceptable level.

	Starting position	End Position	Route
One to One	A B C	D E F	miles miles miles
One to Many	A B C	D, E, F D, E, F D, E, F	2,3,4 miles 2,3,4 miles 2,3,4 miles
Many to Many	A, B, C	D, E, F	2,3,4,2,3,4,2,3,4 miles

Table 4.2 Table showing the different input options within pgRouting

With the datapoints formatted correctly it was possible to get the distances required from pgRouting. This can be done in multiple ways and pgRouting offered a number of algorithms (shortest path Dijkstra, shortest path A*, multiple shortest path Dijkstra, bidirectional A* and others). The most suitable for this project was the Dijkstra algorithm which finds the shortest route through a graph. Within the Dijkstra function there are options that control which information about the selected route is returned to the user for the purposes of a 2SFCA calculation; the only information required is the total distance, and so the PGR_dijkstraCost() function was utilised. Pgr_astar() could have been used too as it allows the routes to be found whilst not calculating the whole graph. On small scale (local or small city level) studies this may be more efficient, but when the whole graph needs to be analysed it is slower than Dijkstra. Pgr_astar does not necessarily return the best route and in certain circumstances it is not as accurate as Dijkstra.

The distances needed to be joined to the other data relating to each datapoint and at this point it was possible to compute the distance decay values. The solution developed offered the option to apply both a linear and a Gaussian distance decay function. These distance decay functions are discussed further in Section 6.5. This step of the processing is skipped if the user has specified no distance decay function. The linear distance decay was used as the standard option and the user had the choice to opt for another parameter. The tool removed any distances that are outside of the specified catchment area, this needed to be completed again as a networked distance can be significantly larger than a Euclidean distance. The demand scores were now summed and the supply:demand ratio produced. Several joins were completed to get the data into the correct format, and the data was scaled using a selected distance decay function. Finally, the second step of the 2SFCA calculation was completed and these scores were made available in a results table. Utilising additional data, the scores were joined to a visualisation table which made visualisation of the results easier for the user. The user was notified that the process was finished and the name of the table they would need to open in QGIS (or similar) to visualise the results.

4.4 Case study

After consultation with Tenovus a case study was completed to understand the effects of adding additional resources or moving current resources within Wales through a number of different operational scenarios. There are many benefits to Tenovus understanding how their services affect people in isolation and alongside current NHS chemotherapy provisions. To do this data was gathered on the current health boards' service locations and capacities as well as potential new sites for the Tenovus mobile unit. Using this data, it was possible to propose new locations within Wales that could improve access to chemotherapy. Several sites have been proposed which could improve overall accessibility and permit a wider understanding of current and modelled chemotherapy provisions.

Tenovus normally uses large supermarket car parks for locating mobile units, and for the purpose of this case study a Geolytix dataset (GEOLYTIX retail points) has been utilised to locate each of the supermarkets in Wales. The data was sorted to show only the large 'megastores' which would have ample parking for the Tenovus unit. Although this case study focusses on supermarkets it is important to acknowledge that other venues may be suitable such as sports grounds, race tracks and that these may provide better alternative sites in some of the rural locations throughout Wales. Tenovus have a great network throughout Wales and could use local knowledge in areas with poor accessibility to find these venues. There were 3 locations of particular interest, these sites were chosen based on their location and size and the fact that they fell into areas of poor accessibility and had the potential to dramatically enhance the accessibility of a region within Wales, other sites had the potential to be of interest but this was an exercise to show Tenovus what was possible with the tool and by highlighting sites that had a larger impact it provided more impact; these have been processed and accessibility maps have been produced (see section 6.2).

4.5 Data

4.5.1 Introduction

This section discusses the data that has been used in the testing of the tool. As with the different FOSS choices there are several data sources that need to be considered and evaluated. For the purposes of developing the tool it was important to understand their respective benefits and limitations and make an informed decision on the datasets with the best fit. Data is central to any result the tool provides, and as such it is important to understand the effects that the data has on the results of the E2SFCA calculation. There are three main data areas, the supply side data (where chemotherapy is provided in Wales), the demand side data (population information or proxies for demand) and the network data (road network for speeds).

For this tool to have any impact at Tenovus, it is important to understand the chemotherapy services that Tenovus provides and those offered by the NHS. Tenovus provided the locations of their services which are spread across three areas in South Wales. Each of the Welsh health boards reports their chemotherapy provisions differently, and this made it difficult to understand where patients were sent for treatment. Freedom of information (FOI) requests were made to each health board to get accurate locations and capacity for each. Despite such attempts at obtaining accurate and timely data on levels of cancer services however, there will always be some ambiguity in the data as patients often travel into England for treatment if the treatment centre is closer to home or offers a specialist type of treatment. The treatment options that are available are continually changing making it important to maintain a temporal consistency in the database; as such these maps will only ever provide a snapshot of accessibility at a particular point in time.

Census data was utilised to understand the population, and potential demand for chemotherapy services in Wales. Variables included in the construction of the Welsh Index of Multiple Deprivation (WIMD) for 2019 provided further insight into how people are affected by cancer in certain locations and this allowed different measures of potential demand to be included in the analysis. It also had the potential to run simulations looking at different areas of interest. Tenovus provided 12 months of treatment data which showed the distances travelled by each user of their services. This data allowed for a more accurate input of the distance parameters within the E2SFCA calculation and generally informed modellers on how far people were travelling for chemotherapy services in Wales, as well as guiding the choice of distance decay parameters used in the FCA implementation models.

There were various commercial and open data sets available but this study focused on two open data sets to compare with one paid service. As the purpose of this project was to utilise free and open-source software it was also important to ensure that the data for the tool was freely available, and to investigate whether such data has a dramatic effect upon the results. In the United Kingdom there are two main sources for open network data, Ordnance Survey Open Roads and Open Street Map (OSM). Google was used as a comparison to see how the different networks perform and see where improvements could be made. Some comparative work was completed on the data gathered to understand any potential positives in the data, and to see the potential effect it may have on the E2SFCA scores produced. A set of routes was calculated using Google Maps, OSM and OS Open Roads and the times and distances produced were compared, the different data sets affected the performance of the tool so it needed to be worthwhile using a more detailed data set which was more computationally heavy. A scale comparison between OA, LSOA and MSOA was completed which investigated how the aggregation of data affected the E2SFCA results (see section 6.4.1).

4.5.2 Supply side data

4.5.2.1 Tenovus data

Tenovus provided current and historical data for all of its mobile units. There are two main units; one specialising in chemotherapy provision and the other providing lymphoedema treatment and education. Historically there was a further unit named the ‘Man Van’ which promoted screening and education regarding prostate cancer in Wales. Currently there are units working in England as well which are not used in Wales. The mobile units are able to move every day and can service large areas. There is potential for the vans to move to two locations in one day as well and the sites included are those which the van goes to on different days of the week. The historical data was used to show researchers from the mobile unit, choir and provision teams within Tenovus the impact of locating such services on the overall levels of accessibility within Wales, whilst the additional static provision data was being sourced. The current locations of the units can be seen in Figure 4-10 and can be used alongside the static supply side locations to provide an overview of potential treatment alternatives in Wales.

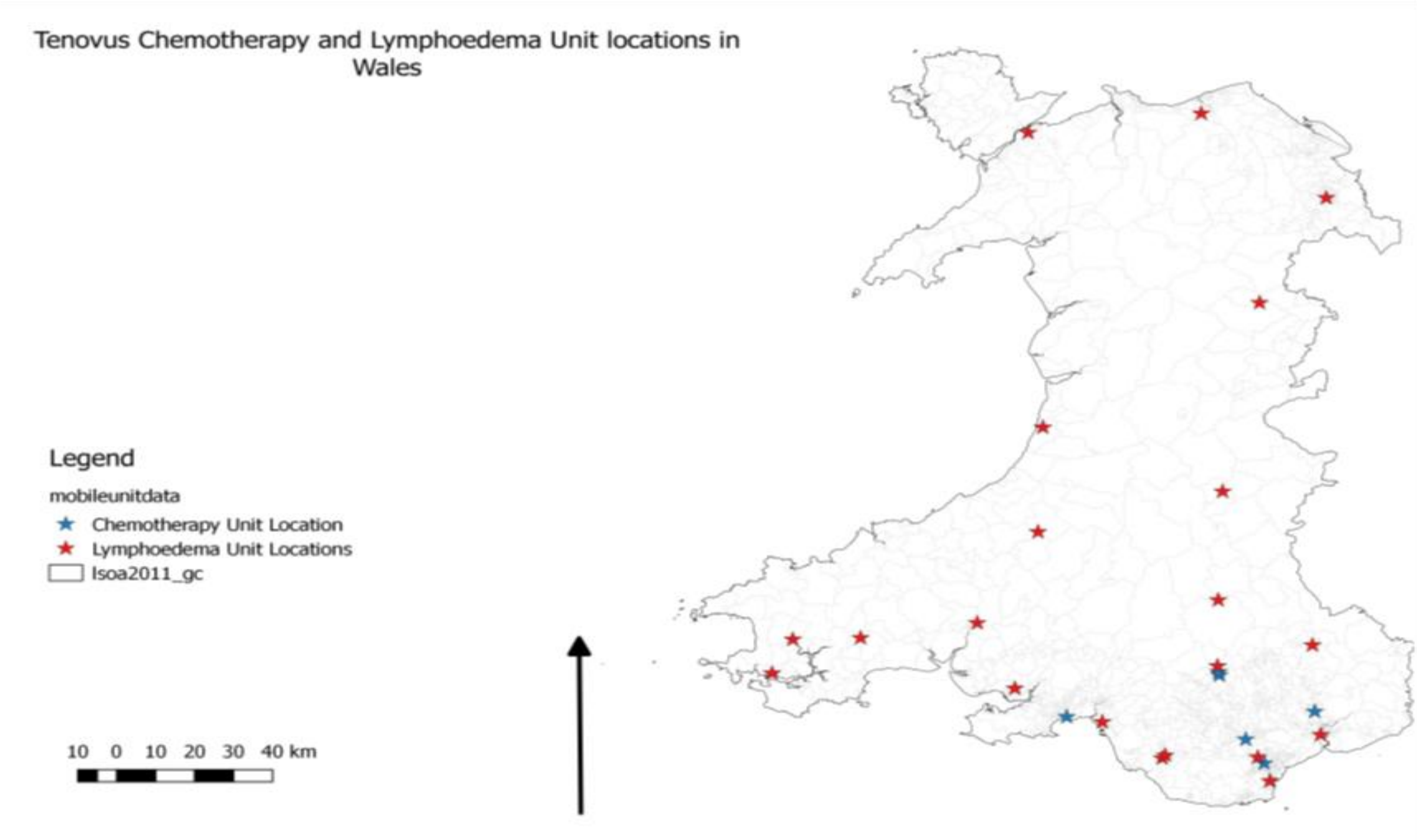


Figure 4-10 Chemo unit is red and lymph unit is blue

4.5.2.2 FOI requests data

Currently in Wales there is no single source of information regarding chemotherapy provision. After numerous attempts to source this a number of FOI requests were made to each of the Welsh health boards (Figure 4-11), to establish the chemotherapy and lymphoedema treatments offered at static locations in Wales. The FOI requests asked the following information:

- At what locations do you deliver lymphoedema treatment?
- At what locations do you deliver lymphoedema education?
- What are the operational days and hours of each location delivering Lymphoedema treatment?
- How many nurses are employed to deliver lymphoedema treatment at each location?
- What is the operational capacity of each location delivering lymphoedema treatment? This should include: The maximum number of individuals that could be treated by location, and the actual number of individuals actually treated by location.

- How many beds or chairs are used for delivering lymphoedema treatment by location?
- What is the average time a patient will spend receiving lymphoedema treatment?
- What is the average distance a patient will travel for lymphoedema treatment at each site?

And for chemotherapy:

- At what locations do you deliver intravenous chemotherapy?
- What are the operational days and hours of each location delivering intravenous chemotherapy?
- How many nurses are employed to deliver chemotherapy at each location?
- What is the operational capacity of each location delivering intravenous chemotherapy? This should include: The maximum number of individuals that could be treated by location and the actual number of individuals treated by location.
- How many beds or chairs are used for delivering intravenous chemotherapy by location?
- What is the average time a patient will spend receiving intravenous chemotherapy?
- What is the average distance a patient will travel for intravenous chemotherapy at each site?

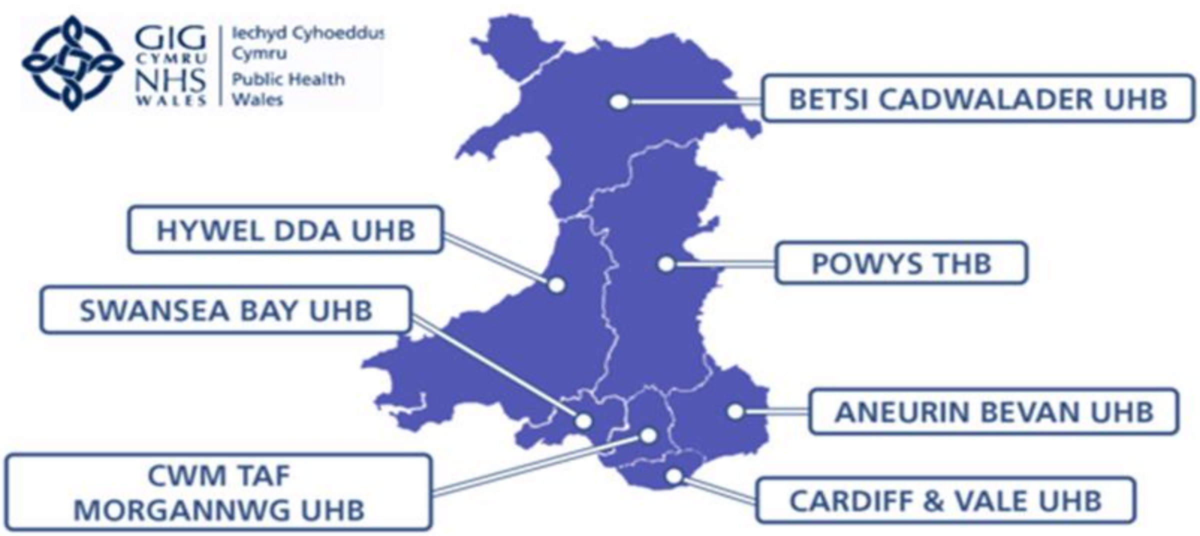


Figure 4-11 Public Health Wales Health board map of Wales[11]

[11] https://phw.nhs.wales/services-and-teams/screening/diabetic-eye-screening-wales/service-user-questions/health-boards-map/

There is also Velindre University NHS Trust which controls the Welsh blood service and Velindre Cancer Centre based in Cardiff. This trust specialises in cancer care and is included here, Velindre is a special case and is located within the Cardiff and Vale UHB to operate the two services mentioned previously.

4.5.2.2.1 Lymphoedema responses

Each health board responded and it was possible to collate the information received in tables 4.3 – 4.7. Velindre does not provide any lymphoedema treatments, but all of the other health boards do to varying extents. Figure 4-12 shows all of the different locations in Wales where it is possible to get lymphoedema treatment or education.

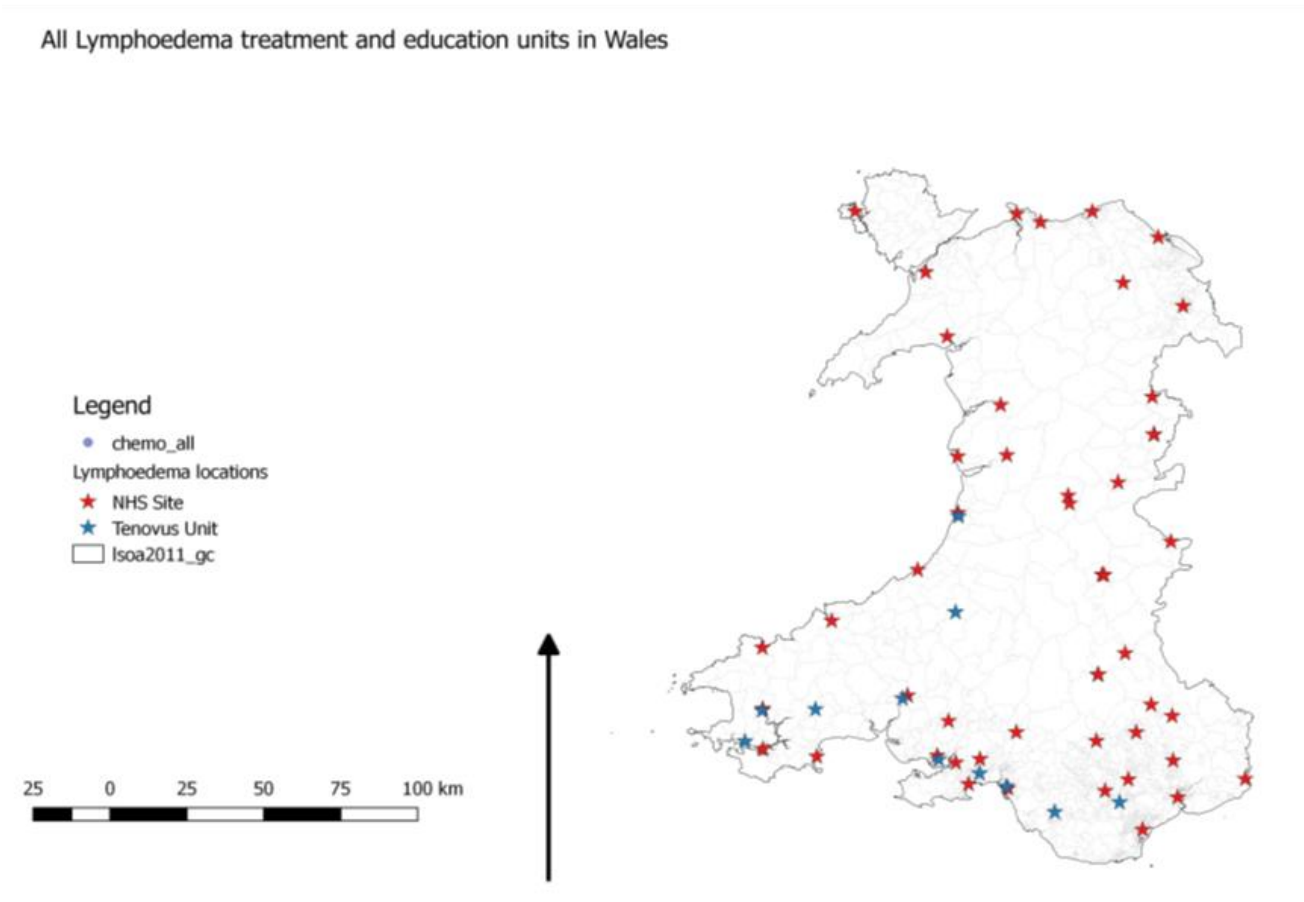


Figure 4-12 Lymphoedema treatment sites in Wales

Location	Opening hours per week	Nurses per week	Capacity	Beds/chairs
Singleton Hospital	42.50	29		8
Neath Port Talbot Hospital	13.50	1.5		
Morriston Hospital	1.00	1		
Gorseinon Hospital Complex Leg Clinic	0.42	1		
Ystradgynlais Leg Club	1.25	1		

Table 4.3 Lymphoedema FOI response from Abertawe Bro Morgannwg

Location	Opening hours per week	Nurses	Capacity	Beds/chairs
St Woolos Hospital	40			6
Ysbyty Ystrad Fawr	8			2
Ysbyty Aneurin Bevan	8			2
Nevill Hall Hospital	8			1
County Hospital (Pontypool)	4			2
Chepstow Community Hospital	4			2

Table 4.4 Lymphoedema FOI response from Aneurin Bevan

Location	Opening hours per week	Nurses	Capacity	Beds/chairs
Department 20, Wrexham Maelor Hospital	42.5	5		3
Satellite Clinic – Flint Community Hospital	40			1
Colwyn Bay Community Hospital	40	2		1
Satellite Clinic – Llandudno Hospital	40			1
Lymphoedema Clinic, Ty Elan Surgery Rhyl	40	2		1
Ruthin Community Hospital	40	2		1
Eryri Community Hospital, Caernarfon	40	2		2
Satellite Clinic – Penrhos Stanley Hospital	40			1
Dolgellau Community Hospital	40	2		1
Satellite Clinic – Alltwen Community Hospital	40			1
Satellite Clinic –Tywyn Community Hospital	40			1

Table 4.5 Lymphoedema FOI response from Betsi Cadwaldr

Location	Opening hours per week	Nurses	Capacity per week	Beds/chairs per week
Dewi Sant Health Park	16	2	40	2
Keir Hardie University Health Park	40	3.5	146	3.2

Table 4.6 Lymphoedema FOI response from Cwm Taf

Location	Opening hours per week	Nurses	Capacity	Beds/chairs
Withybush Hospital	35.5			2
Satellite Clinic – Tenby Cottage Hospital	8			2
Satellite Clinic – South Pembrokeshire Hospital	8			2
Satellite Clinic – Fishguard Health Centre	8			2
Ty Geraint, Bronglais Hospital	40			2
Satellite Clinic – Aberaeron	16			2
Satellite Clinic – Cardigan Community Hospital	8			2
Lymphoedema Clinic, Prince Philip Hospital	40			2
Satellite Clinic – Ty Cymorth	24			2
Satellite Clinic – Glangwili Hospital	8			2
Satellite Clinic – Cross Hands Health Centre	4			2

Table 4.7 Lymphoedema FOI response from Hywel Dda

4.5.2.2.2 Chemotherapy responses

Figure 4-13 is a map showing the chemotherapy locations in Wales. Cwm Taf and Powys do not provide these services, and most of the people living within these areas travel outside their Board areas for treatment. Tables 4.8 to 4.12 detail the responses received from each health board and show the raw data that is being worked with in this study.

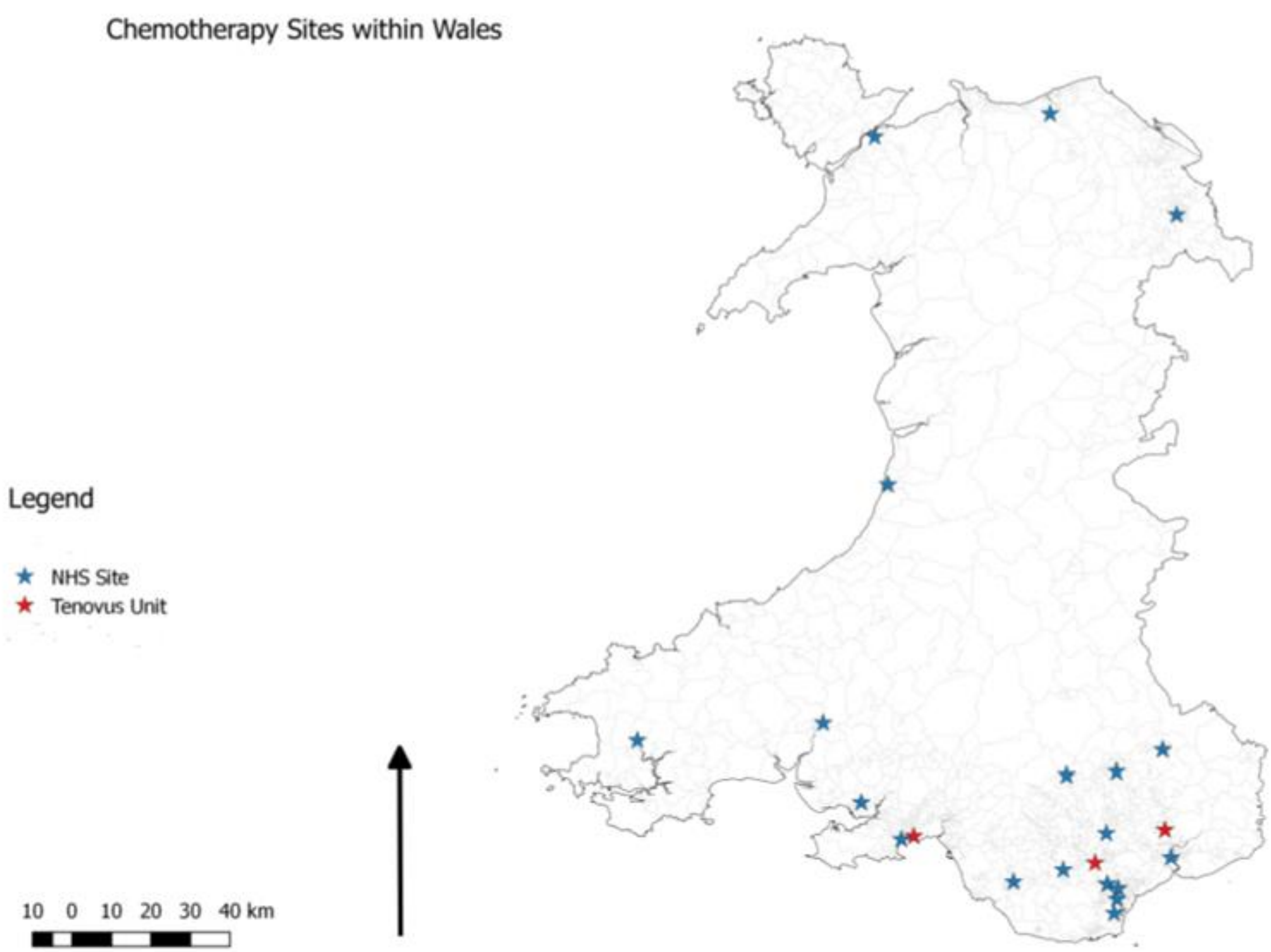


Figure 4-13 Chemotherapy location in Wales

Location	Opening hours per week	Nurses WTE	Capacity	Beds/chairs
Ysbyty Maelor Wrecsam	42.5	9.15	30	11
Ysbyty Glan Clwyd	42.5	9.86	40	21
Ysbyty Gwynedd	42.5	7.31	30	20

Table 4.8 Lymphoedema FOI response from Betsi Cadwaldr

Location	Opening hours per week	Nurses WTE	Capacity	Beds/chairs
Singleton Hospital	50	45.39	55 (avg per day)	36

Table 4.9 Chemotherapy FOI response from Abertawe Bro Morgannwg

Location	Opening hours per week	Nurses WTE	Capacity	Beds/chairs
Royal Gwent Hospital	40	9		26
Nevill Hall Hospital	32	5		11

Table 4.10 Chemotherapy FOI response from Aneurin Bevan

Location	Opening hours per week	Nurses WTE	Capacity	Beds/chairs
Bronglais General Hospital	40	2.76		6
Glangwili General Hospital	40	4.8		8
Prince Philip Hospital	40	3.2		6
Withybush General Hospital	40	4.4		8

Table 4.11 Chemotherapy FOI response from Hywel Dda

Location	Opening hours per week	Nurses WTE	Capacity	Beds/chairs
Chemo inpatient unit Velindre Cancer Centre	45	4	6	14
Chemo day unit Velindre	45	6	30	12
Rhosyn day unit Velindre	42.5	5	25	9

Prince Charles Hospital	40	5	24	10
Royal Glamorgan Hospital	30	2	12	3
Nevill Hall Hospital	40	5	24	10
Princess of Wales Hospital	25	2	12	3
Ysbyty Ystrad Fawr	15	2	6	4
St Davids	27.5	2	12	4
Ysbyty Aneurin	15	2	6	3

Table 4.12 Chemotherapy FOI response from Velindre

The results of the FOI request show that there are large areas in Wales with poor levels of access to chemotherapy services. For example, those patients who live in Powys are forced to travel quite large distances for treatment, whether in England or in one of the locations within Wales that offer different types of treatment. It stands out that most of the locations of current provision are in the areas with higher populations. However, even if potential users of the service reside close to these locations, they may not provide for every type of chemotherapy service and ultimately individuals may need to travel further to obtain treatments and inevitably such patients will share these services with other potential users of these services.

Using the FOI request data, it was possible to use a more advanced supply value. One of the ways this was done was to use the number of hours the site is open multiplied by the number of beds/chairs they have available for treatment. This supply-side value allowed for a more accurate understanding of the actual supply at each location as opposed to using the number of beds/chairs or opening hours on their own (section 6.2). It was an important consideration to have all of the data reported at this level as most of the online sources did not show the data at a small enough scale to enable it to be used accurately. This data was used in the case study (Section 6.2) to help the Tenovus research team demonstrate current levels of service provision and understand where the gaps in provision occur. The case study utilised the major retail sites in Wales and suggested locations for additional chemotherapy provision, this allowed Tenovus to understand the impact that they are currently having on chemotherapy provision in Wales and the ways in which they could potentially increase the impact of their services.

Figure 4-14 shows the level of accessibility for people within a 35 km catchment area, and it is possible to see the gaps in provision. The 35km distance was used as analysis of the Tenovus utilisation dataset and shows that the vast majority of patients are within a 35km catchment area (section 4.3.3). It also highlighted that those in areas with a large amount of provision did not necessarily have better access to those services. For example, in and around the Cardiff area there are a large number of sites but only small pockets within the area have a good access to these services because of the levels of potential demand on them.

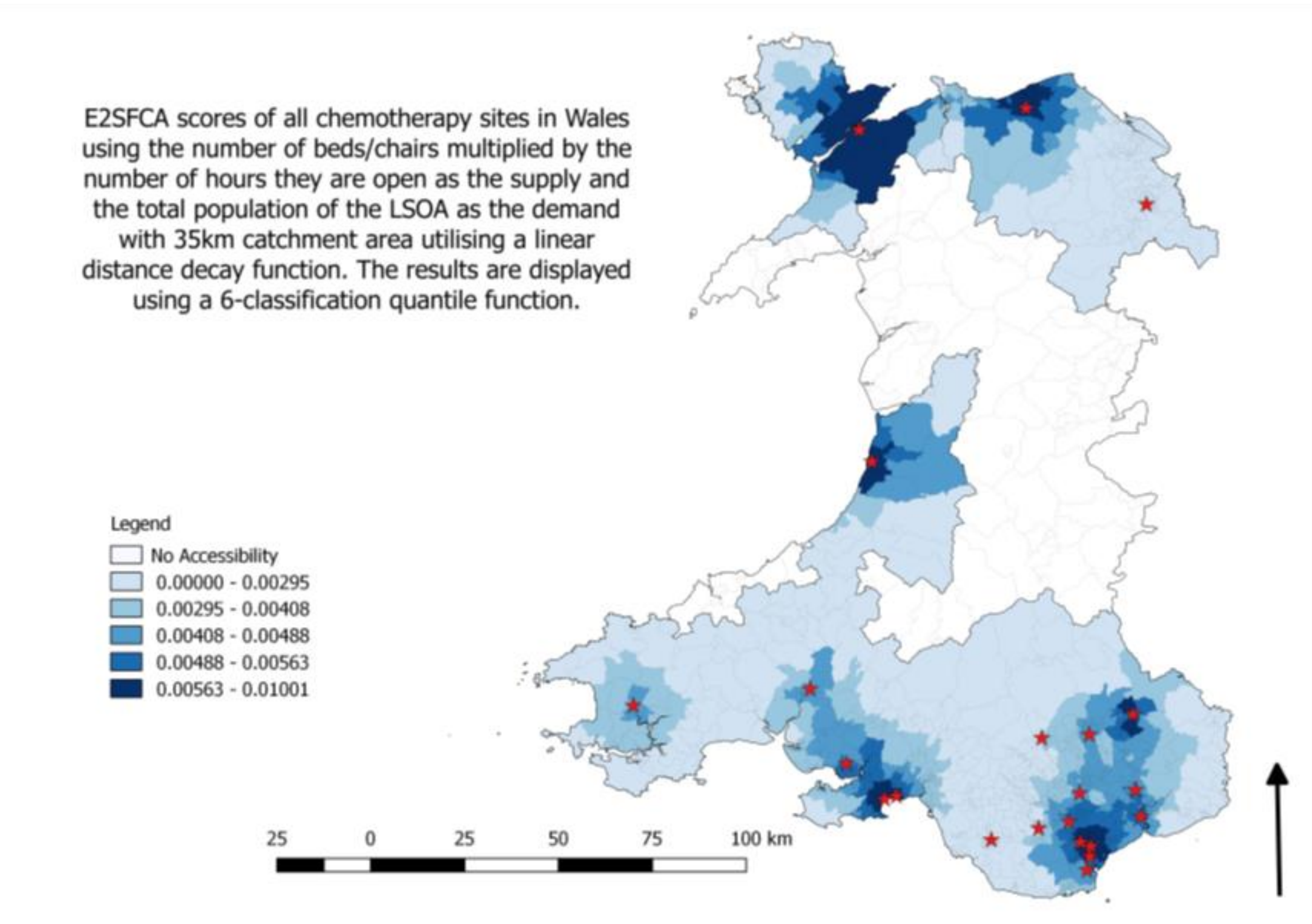


Figure 4-14 E2SFCA results chemotherapy sites in Wales as supply with LSOA total population as demand

4.5.3 Demand side data

4.5.3.1 Census data

The data used to understand the demand for chemotherapy services in Wales was sourced from the 2011 Census of Population, and variables that are used to calculate the various domains of the Welsh Index of Multiple Deprivation (WIMD). Lower Layer Super Output Area (LSOA) is a type of geographical unit used in England and Wales to display data at local level containing at least 1,200 households. Output Area (OA) are similar to LSOAs but they are even smaller areas that are directly linked to census data. They are required to have a minimum of 40 households. Middle Layer Super Output Area (MSOA) is the largest area looked at in this section; they contain at least 6,000 households. There is a considerable amount of data released at this spatial resolution, and it is important to understand how the level of spatial aggregation impacts on E2SFCA results. LSOA level data was predominantly used as it was the smallest area that much of the meaningful data could be obtained because of confidentiality constraints.

It is important to note that chemotherapy is not something in which someone normally has a choice, demand scores normally relate to something which is a choice (screening, exercise, etc.), and chemotherapy normally provides a critical service required by the population. To account for this it is important to look at more sensitive data which can more directly account for those who may require chemotherapy. The WIMD combines many different metrics to present an index score to evaluate the wellbeing of an area. To make this score there are metrics assigned from health, income, employment, education, housing, community safety, access to services, and the physical environment. Within the different metrics used to establish the wellbeing of an area, one of particular interest is the cancer incidence (indirectly age-sex standardised) (number by 100,000) at LSOA level. This figure was used to create a measure of potential demand to compare with the total population score provided by the census. Using a combination of the census and the WIMD it was possible to focus the data on particular parts of the population. For example, it may be of interest to investigate 60+ year old males when investigating the accessibility of prostate cancer services.

Median population weighted centroids were used to represent a single geographical location for each MSOA (Figure 4-15, LSOA) (Figure 4-16) and OA (Figure 4-17). Population weighted centroids are singular points within output areas that represent the distribution of the population in each output area at the last census (available from www.data.gov.uk[12]). These are commonly used in this type of study as they are easy to understand and represent the best estimation of where the population will potentially be travelling from. The benefit of using a centroid is that it allows for a single point to be identified as opposed to a polygon which would better represent the area of an area but would be much more troublesome to compute and visualise. Total population was also used from the 2011 census data which gives a value for each LSOA.

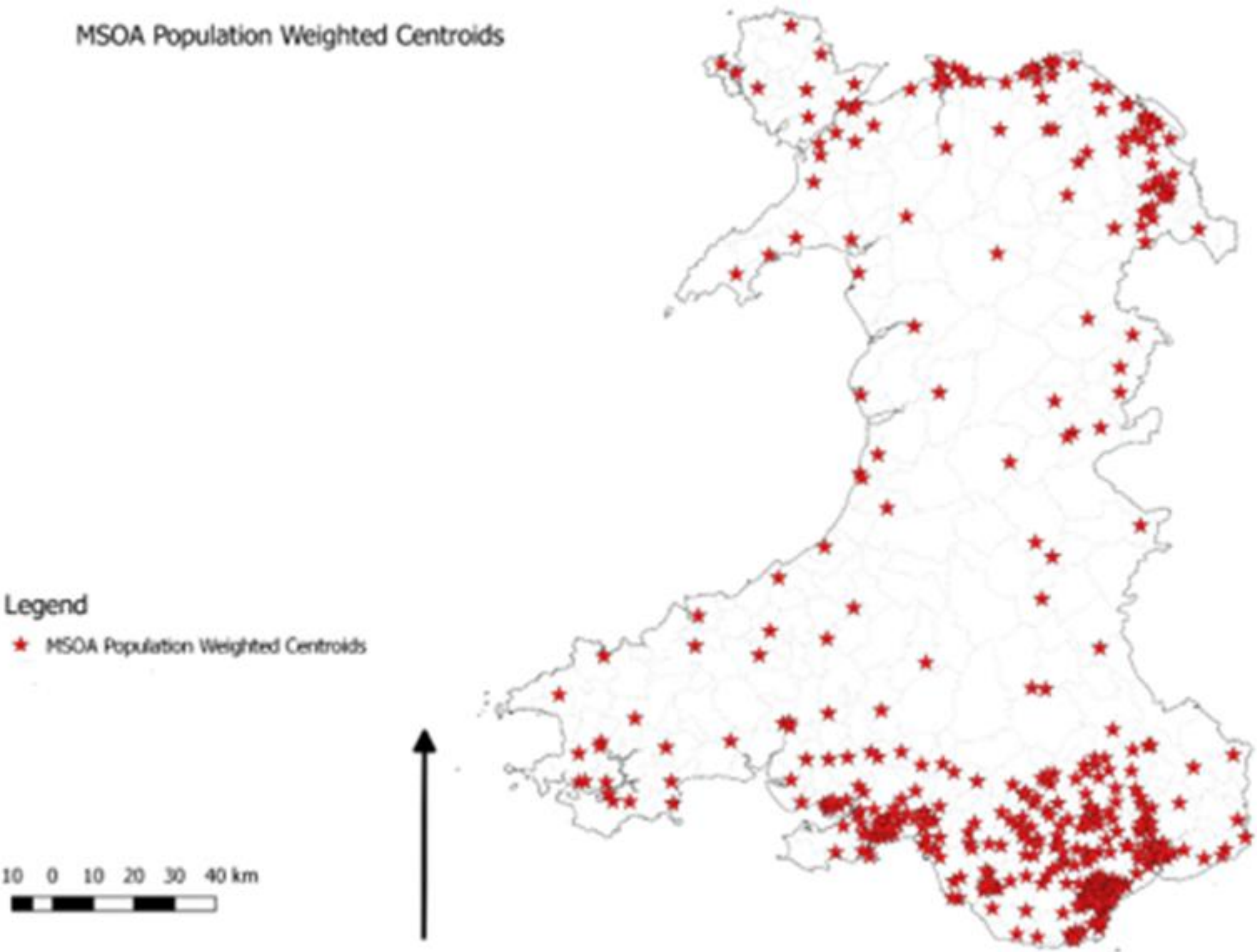


Figure 4-15 MSOA centroids

[12] https://data.gov.uk/dataset/a40f54f7-b123-4185-952f-da90c56b0564/lower-layer-super-output-areas-december-2011-population-weighted-centroids

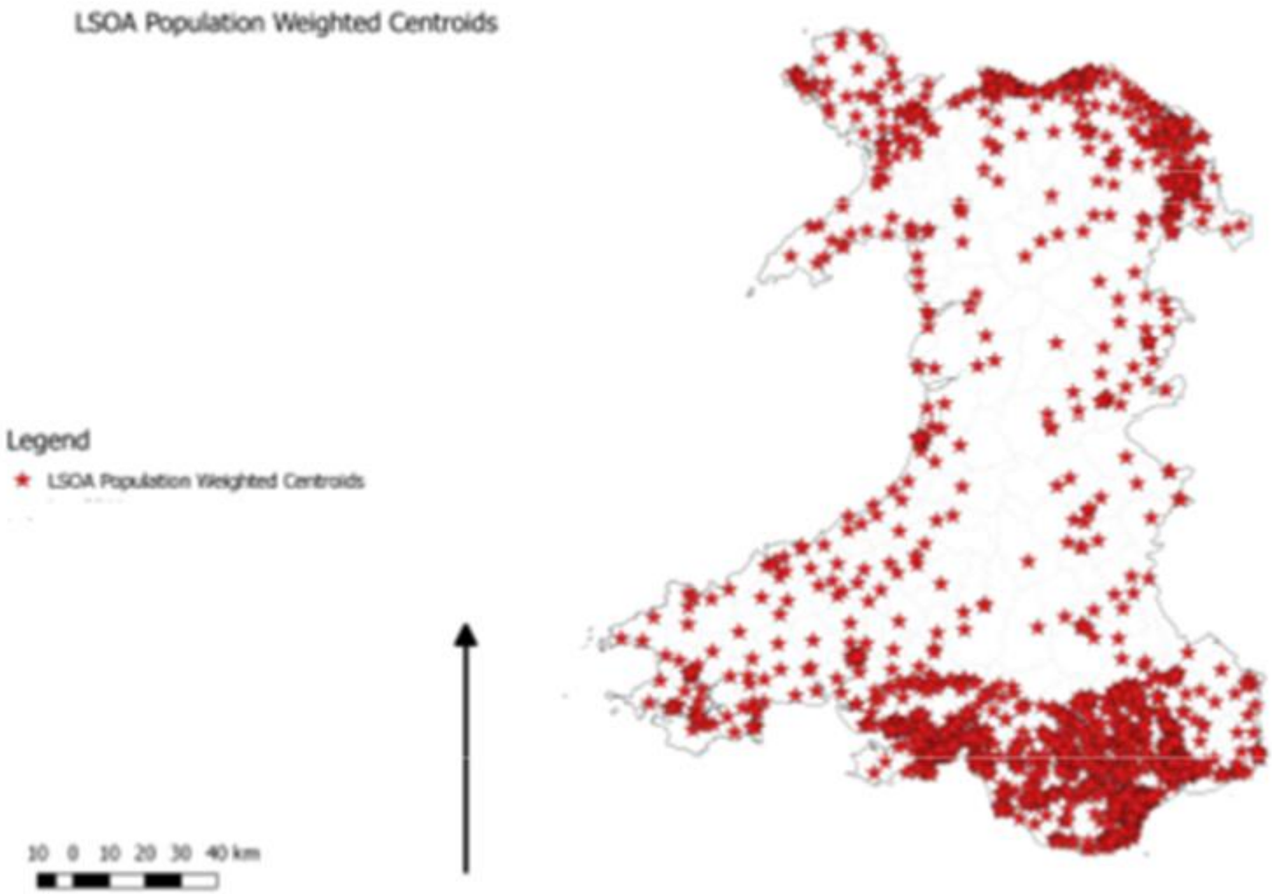


Figure 4-16 LSOA Centroids

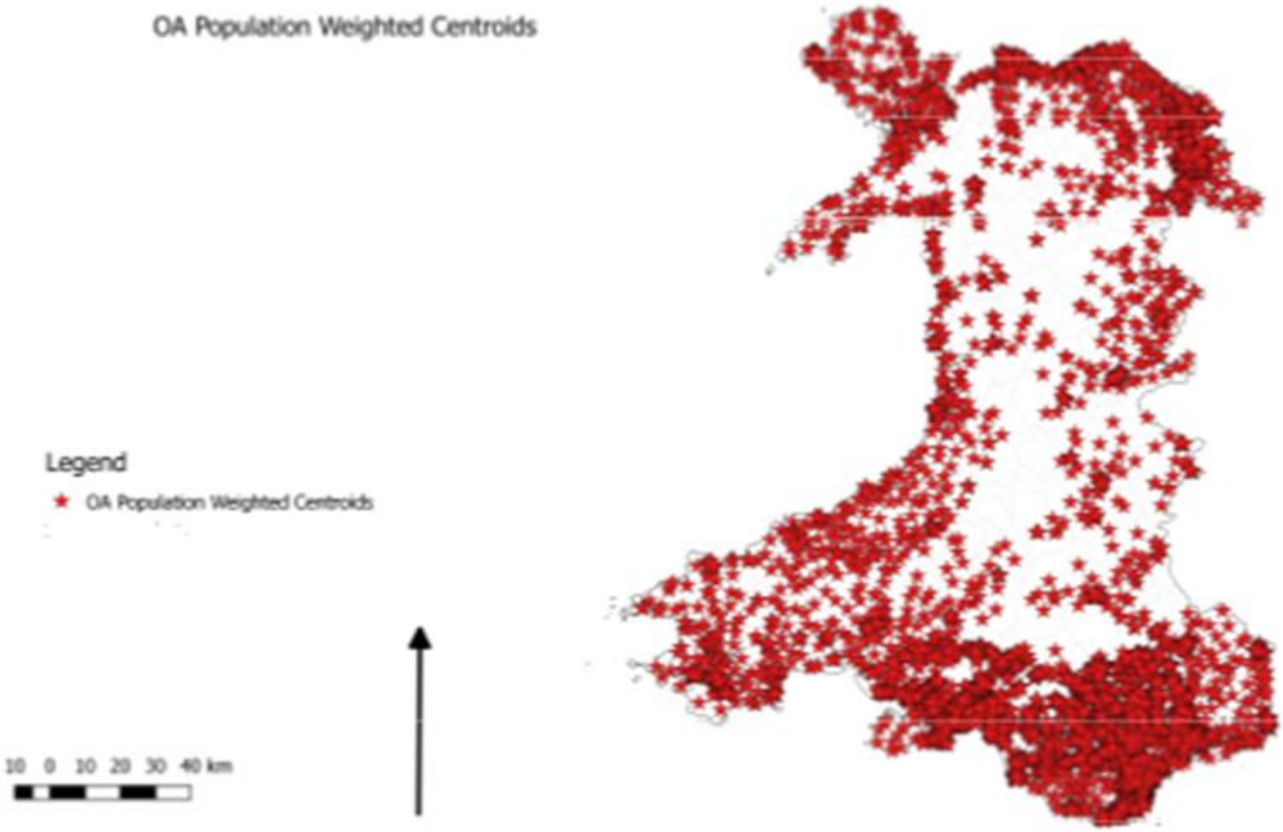


Figure 4-17 OA Centroids

4.5.3.2 Distances travelled between Tenovus users and Tenovus mobile units

In order to gain an understanding of the actual utilisation behaviour of patients Tenovus provided a dataset which showed the start and end journey of users of their services over the course of one year. The data showed the type of activity the unit is performing (Education, Lymphoedema or Chemotherapy), the cancer diagnosis of the user, the postcode of the user, the date of the treatment and the location of the unit. This section explored the distances travelled by users of these services, and broke down the totals in order to gain an insight into how far a typical user of the services would travel, and guide the parameters used in the FCA models. Understanding this data assisted in the size of catchments used going forward and potentially allowed for more accurate 2SFCA scores. Normally the catchment area is based on a 'best-guess' or anecdotal evidence so having access to actual user data increased the accuracy of the results and gave Tenovus a better understanding of their users. This data was not available from the static sites as requested in the FOI request.

4.5.3.3 Total Distances travelled

This section looks at the distance travelled by Tenovus users, Tenovus provided the home postcode for every user of their services in a 12-month period. Using the home postcode of the user and the location of the mobile unit, a distance for each treatment was recorded. This data is rarely available to researchers but gave an insight into the distances that people travel for treatment, and gaining an understanding of this data allowed for the E2SFCA parameters to be guided by real data instead of anecdotal data or guesswork. Table 4.13 shows that the average user of the mobile units was travelling about 10km by car; this should take a maximum of 15 minutes each way, and most of the locations have ample car parking. The longest journey was a 4-hour round trip and the shortest was 200 metres. As the histograms presented in Figures 4-18 and 4-19 show, the vast majority of users travelled under 20kms to receive their treatment.

	Euclidean (metres)	**Network (metres)**	
Mean journey	9,133.34	11,437.67	
Median Journey	7,209.71	9,252.02	
Longest journey	132,987.29	154,258.64	This journey is from Haverfordwest (SA61) to Cardiff (CF24). Google estimates that the journey would take around 2 hours in a car dependent on the time of day.
Shortest journey	200.55	358.58	This journey is to the Brecon location and there are several users that have to travel less than 1 km to access the unit.

Table 4.13 Total distance travelled to a mobile unit

Figures 4-19 and 4-20 show histograms of the data and enable a clear view of the journeys being completed by Tenovus users. The data shows that the vast majority of use is within a short distance of the site, and that the data is Gaussian in nature in as much as access remains stable for a short distance then drops dramatically similarly to a Gaussian curve. A Gaussian curve is depicted in figure 4-18 and although this data only represents half of the Gaussian curve it is possible to see the similarities where the scores remain relatively high then quickly decrease before leveling off again in a long tail.

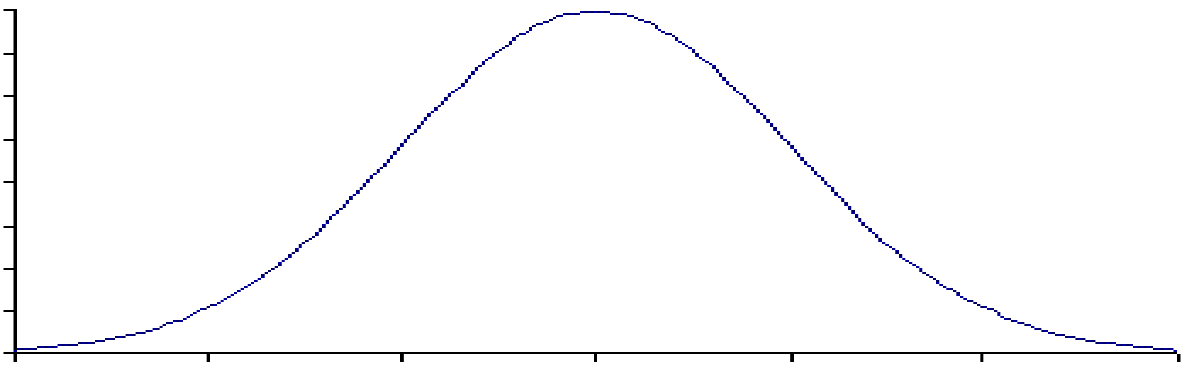


Figure 4-18 example of a Gaussian curve

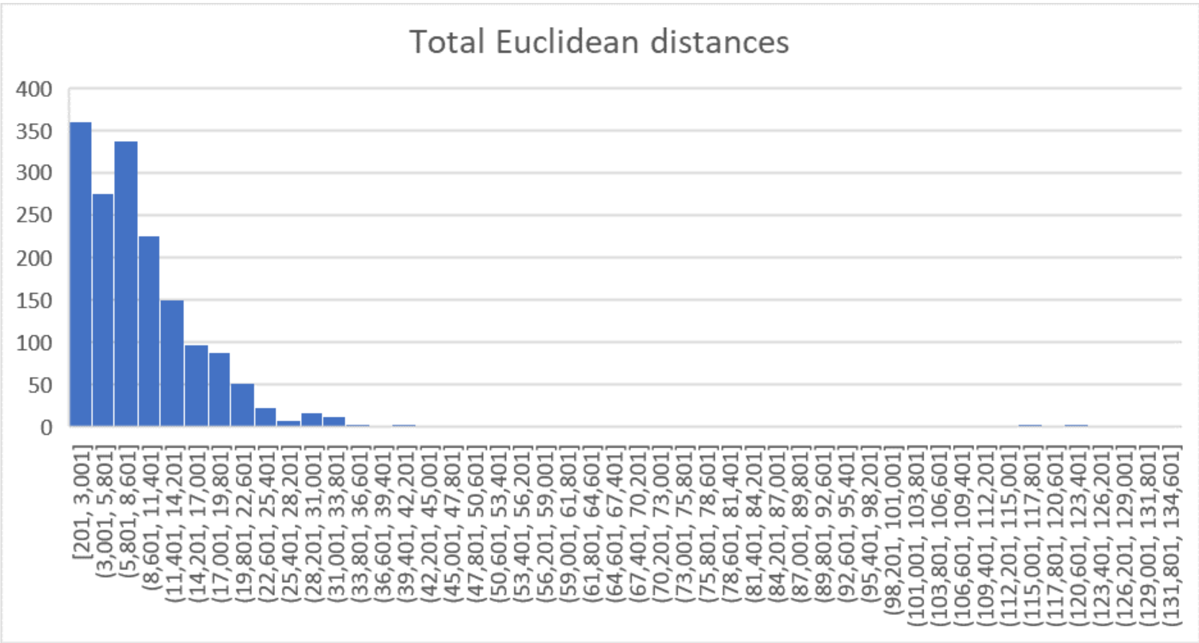


Figure 4-19 Total Euclidean distance where the x axis is the number of journeys and the y axis is a range of distances travelled in metres

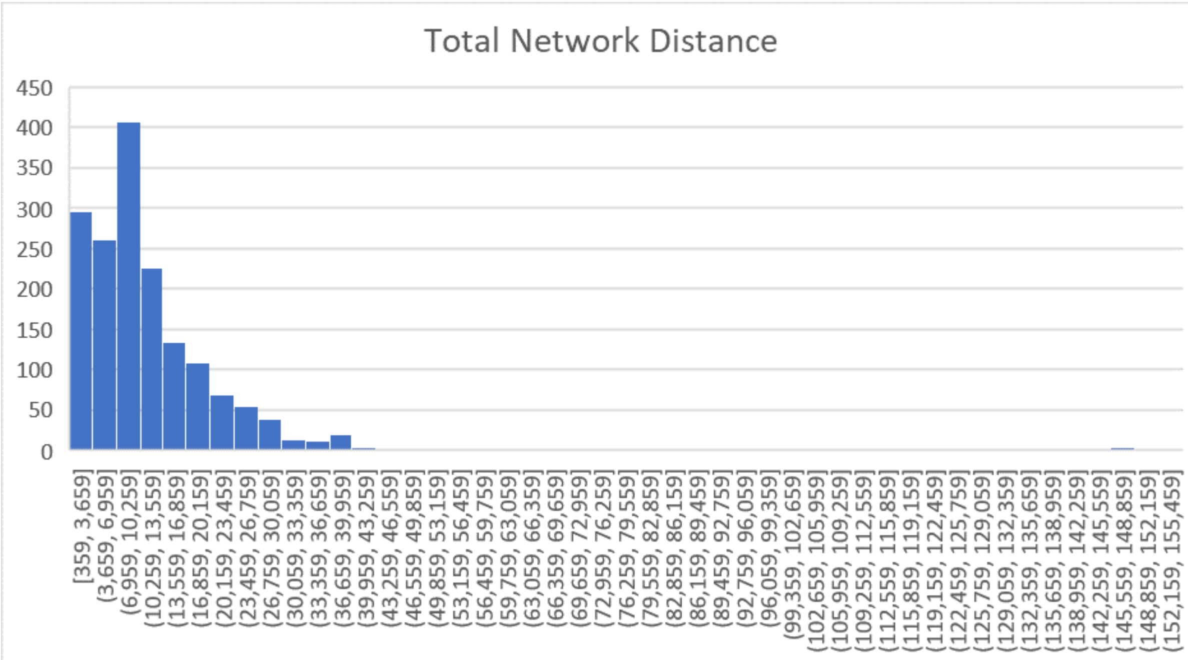


Figure 4-20 Total network distance where the x axis is the number of journeys and the y axis is a range of distances travelled in metres

4.5.3.4 Distance travelled for chemotherapy

Table 4.14 shows that the longest journey for any user was for a chemotherapy session, but the histograms show that this was unusual, with the majority of users travelling less than 20km to the mobile unit. Table 4.15 shows that breast cancer treatment accounts for around half of all treatments provided on the chemotherapy unit, with blood cancers next.

	Euclidean (metres)	**Network (metres)**	
Mean journey	9,896.35	12,443.51	
Median Journey	7,568.30	9,682.11	
Longest journey	132,987.29	154,258.64	This journey is from Haverfordwest (SA61) to Cardiff (CF24). Google estimates that the journey would take around 2 hours in a car dependent on the time of day
Shortest journey	389.29	600.65	This journey is between two locations in Cwmbran

Table 4.14 Distance travelled to the chemotherapy mobile unit

Cancer type	**Treatments provided**
Breast	467
Haematological	230
Colorectal	179
Gynaecological	6
Lung	9
Urological	13
Upper GI	30

Table 4.15 Chemotherapy treatments provided

Figures 4-21 and 4-22 show histograms of the Euclidean and Network distances travelled by the chemotherapy users of Tenovus, the data is again Gaussian in nature.

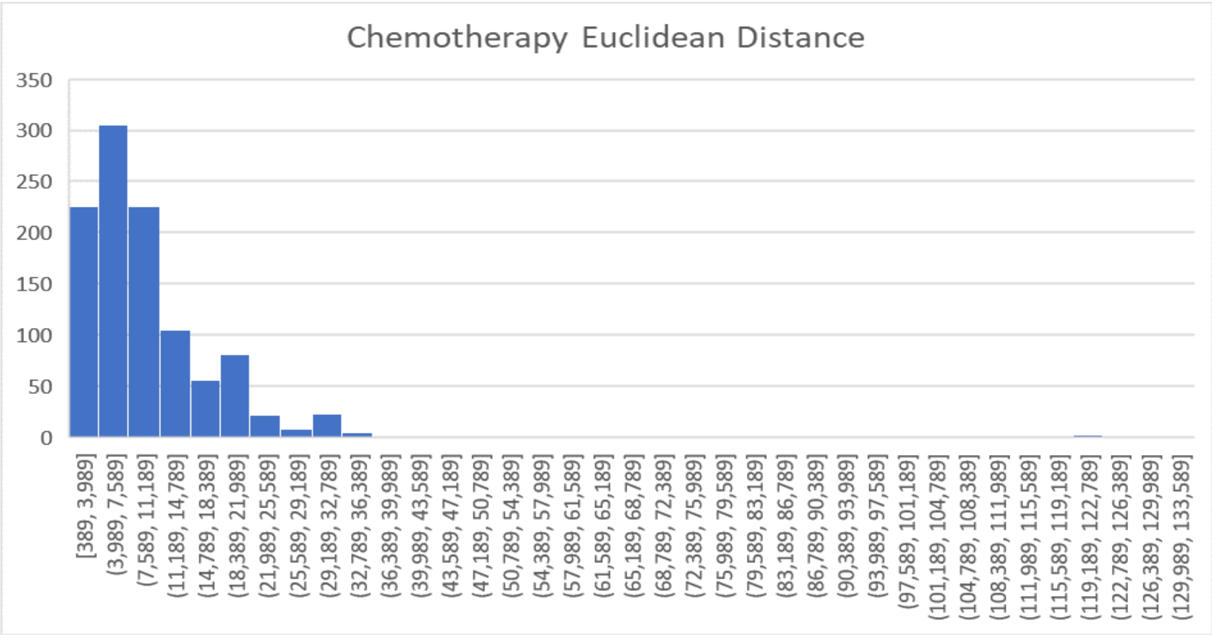


Figure 4-191 Euclidean Chemotherapy distance where the x axis is the number of journeys and the y axis is a range of distances travelled in metres

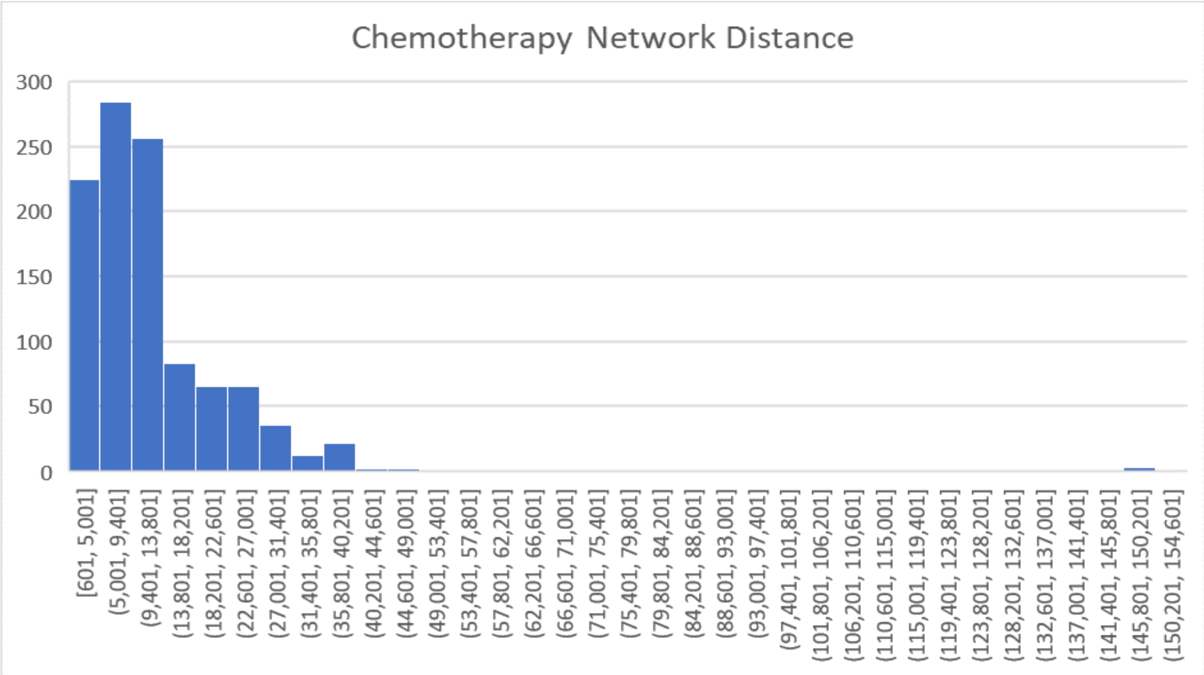


Figure 4-202 Network Chemotherapy distance where the x axis is the number of journeys and the y axis is a range of distances travelled in metres

4.5.3.5 Distance travelled for Lymphoedema

Table 4.16 shows the lymphoedema unit has much lower travel times than the chemotherapy unit. The majority of users are within 11km, with a maximum journey of 43km. Lymphoedema treatment is much more commonly available, and the need to travel for the service is reduced. Table 4.17 highlights that breast cancer is again the cancer affecting the majority of the users of their services.

	Euclidean (metres)	**Network (metres)**	
Mean journey	7,319.96	8,990.134	
Median Journey	6,651.28	8,453.12	
Longest journey	38,740.97	43,247.72	This journey is between Llandovery and Carmarthen and should take between 45 minutes and an hour to complete each way.
Shortest journey	200.55	358.58	This journey is to the Brecon location and there are several users that have to travel less than 1km to access the unit.

Table 4.16 Distance travelled to the lymphoedema mobile unit

Cancer Type	**Treatments provided**
Breast	295
Haematological	8
Colorectal	7
Gynaecological	54
Urological	17
Skin	20
Sarcoma	6
Head and Neck	8

Table 4.17 Lymphoedema treatments provided

Figures 4-23 and 4-24 show histograms of the distances travelled by users of the lymphoedema treatment services offered by Tenovus. They are slightly different in nature to the those shown in figures 4-23 and 4-24 and are more linear in their decline.

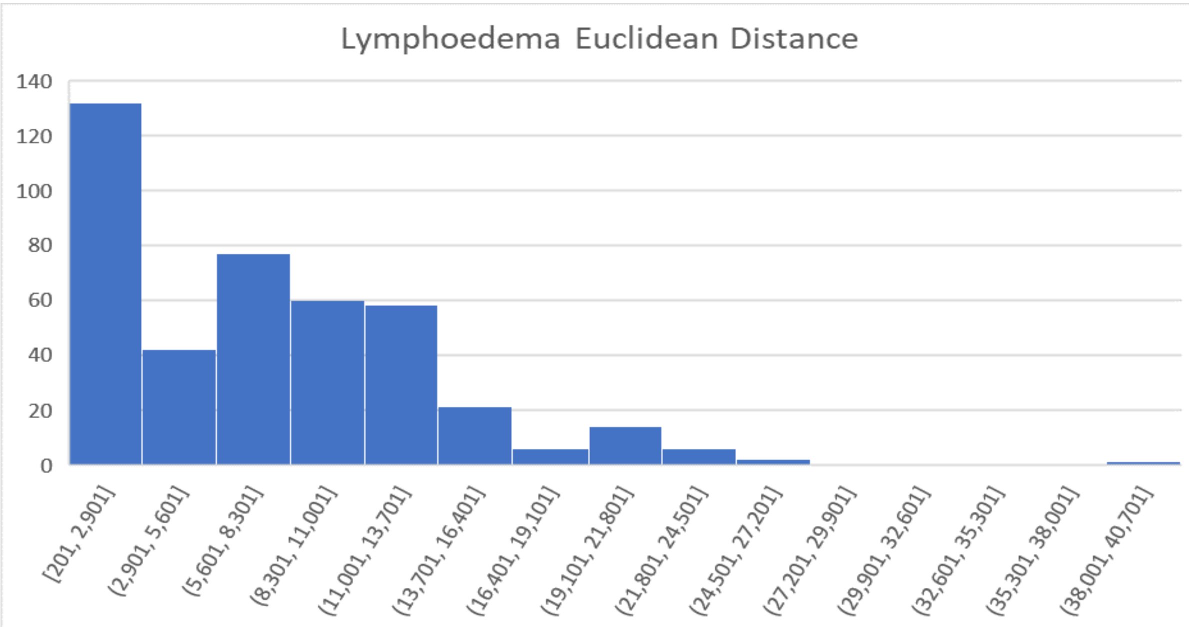


Figure 4-213 Euclidean distance to Lymphoedema Sites where the x axis is the number of journeys and the y axis is a range of distances travelled in metres

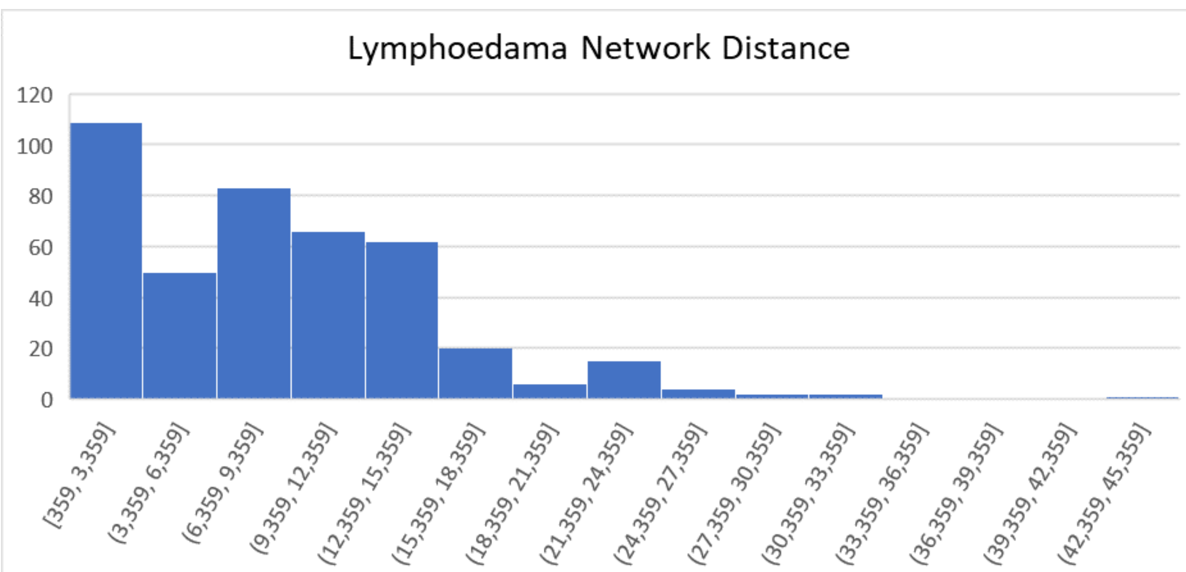


Figure 4-224 Network distance to Lymphoedema Sites where the x axis is the number of journeys and the y axis is a range of distances travelled in metres

4.5.3.6 Distance travelled for education

Educational services are provided on the lymphoedema unit; they are needed to help people with lymphoedema understand their condition and how to best treat it. The vast majority of users travelled less than 15km to use the service. Table 4.18 shows mean and median journeys as well as the longest and shortest journey. Table 4.19 shows that the vast number of users of this service have breast cancer.

	Euclidean (metres)	**Network (metres)**	
Mean journey	12,708.19	14,996.22	
Median Journey	11,423.86	14,027.73	
Longest journey	31,072.93	35,815.11	This journey is from Lampeter to Carmarthen and it should take between 45 minutes and an hour each way.
Shortest journey	1,103.96	1,317.54	This journey took around 5 minutes in a car between two points in Carmarthen.

Table 4.18 Distance travelled to the education mobile unit

Cancer Type	**Treatments provided**
Breast	14
Haematological	1
Sarcoma	1

Table 4.19 Cancer types serviced by the education mobile unit

4.5.4 Network data

Euclidean distance is a good guide but it does not allow for a more detailed understanding of how people travel between two points. The most accurate way for this tool to be utilised is to use a road network and take into account how people would drive from one location to another. For this to happen a road network data set is needed, there are several available commercially; these can include speed limits and be as detailed as required. This project is developing a FOSS solution and for this open data needs to be utilised, in the UK there are two main open data sets, OSM and OS Open Roads.

The OSM initiative started in 2004 and uses local collaborators to map the world. In previous years it was not considered the most accurate (Mooney, Corcoran and Winstanley, 2010) and in some parts of the world it is still not fully complete (Sehra, Singh and Rai, 2017). In the UK, a large amount of OSM data is available, but to date there have been a limited number of studies that considered its utility in FCA studies (Frew et al, 2017). Most roads have been tagged with individual speed limits and it is also possible to use it with a focus on public transport or cycling, but that is not required for this particular application. Although there are a variety of tools available to create a network from OSM data (osm2po, osm2pgsql) these can be difficult and could be daunting to a non-expert as they involve using the computer terminal instead of a graphical user interface.

OS Open Roads is offered on an open government licence. It does not contain as much information as the OSM dataset but it benefits from being in shapefile format, so it is quite easy for a non-expert to upload to a spatial database. It is possible to buy the more detailed road network from OS, and it is also an option to utilise the additional data provided from OS and their partners but these are all paid services.

The main downfall for OS Open Roads is how the roads are divided into seven different categories which all need to be assigned a road speed. This means that every dual carriage way in Wales has the same speed attached. This does not provide any flexibility, and potentially will result in routes that are less accurate than those in OSM.

4.6 Process for visualisation of FCA results

The tool created two files which could be visualised so that the results are easy to interpret and share. To assist the user in this there is a file with all of the results in a standard database table which simply shows the FCA score, the location and unique identifier. This information is then joined to the LSOA boundaries map so that it can be easily visualised. For the purposes of this project, it is used with QGIS which provides a full FOSS desktop GIS solution and could prove to be useful for researchers in Tenovus for a wider range of projects. To show the map in QGIS several simple commands are required and described in the documentation, which allows them to share a map between offices and present the map to others inside and outside the Tenovus organisation. Although QGIS is recommended for use with this tool it should work equally well with any other GIS that PostGRESQL can connect with.

4.7 Documentation

The use of this tool is dependent on the quality of the documentation provided; this is particularly important as the people using it have differing capabilities with computer systems. During the UX study it was established that the user instructions were as important as the program itself, and having clear and well-designed documentation enabled users to understand what they were doing and why. During the UX study it became clear that there would be a need for different documentation formats for different types of user, and that a full technical document would be needed to install and assist in the updating of the tool; a simple set of instructions which the end user can follow to get the 2SFCA results that they need, and some documentation about what the results mean and how they can be interpreted.

The super user documentation assists in the entire set up of the tool. It discusses where to get the data that is required and how to organise it alongside the installation of the appropriate software and the versions required. The differences between using different network data are shown and the correct formats are presented. It is important for an administrator to add new data to the database at certain points and there are simple step by step instructions on all of the tasks that the administrator will be required to perform.

The user documentation reads more as instructions with screen shots. This allows the user to see where they are and understand how to fix what they may have done wrong. Taking this on a screen by screen basis allows the user to get real help from the document, and by highlighting what needs to be changed the user should be able to stay on track. There are only a few inputs required, and the user is able to know the process after 2 or 3 attempts.

The final document is needed to explain what 2SFCA is and what the results mean, and is relatively easy to understand. It is important for the user to see the benefit of this analysis and compare it to other methods in use. There is a further explanation of how to read the results on the map and why these results are the way they are.

4.8 Chapter Summary

This chapter discusses the process involved in the research and development of the second iteration of the 2SFCA tool, the different FOSS options available for a tool such as this and the advantages of certain pieces of software. There is a detailed description of the algorithm used to compute the different E2SFCA scores and justifications for the choices made regarding the routing algorithm used. A description of the case study completed for Tenovus looking at modelling the implications of different locations for their mobile unit within Wales is discussed. For this project to be successful a large amount of accurate data is required to provide any meaningful results. This section described the different data and how it has been sourced and formatted.

As there are so many different ways that this project could have been completed, it is important to understand the current FOSS options available for the development of a project like this. By evaluating many of the options in detail it was possible to present an overview of the FOSS GIS systems available and their respective strengths and weaknesses in relation to the project aims outlined in Chapter 1.

There are many routing options available; if this project was intended for a sole user with a small static dataset it may have been possible to use a more efficient algorithm to compute the routes. To keep the flexibility of the tool and ensure that there are no large loading times the Dijkstra algorithm was used with the cost only function which reports a single summed total distance or time for the route.

A case study was completed investigating the potential changes in accessibility to cancer services in Wales by moving the Tenovus mobile unit locations. Several sites were identified, and using the 2SFCA tool the accessibility maps were computed and described. This case study was not intended to be exhaustive but to show researchers within Tenovus the type of results that could be gained by using the tool.

The data used in this project is incredibly important, and understanding the implications of data quality on the resulting patterns is important for assessing the validity of the results. There is a lot of data that could be purchased, as well as private data within organisations, which would provide insight into accessibility to chemotherapy services within Wales, but this project is limited to open data, and the data provided by Tenovus and several FOI requests. The data is as important to achieving meaningful results as are the algorithms which run the tool because the scale and abstract nature of the data has an important impact on findings.

This chapter has described many of the processes involved in the research and development of the tool, and this has enabled some useful scientific work to be completed alongside the development of a tool. By using these insights, it has been possible to get the best mix of performance and usability in the final iteration of the tool.

Chapter 5 Usability Studies

5.1 Introduction

This is a stand-alone chapter; it discusses the rationale behind undertaking user centred research by conducting a review of the types of techniques adopted in usability studies in the academic literature, and describing the two studies conducted with researchers based at Tenovus offices. This section describes the methods employed and provides the principal findings from this exercise.

One of the fundamental goals of this project was to develop a tool which is usable by non-experts, and to achieve this the user had to be at the centre of the design. Working with Tenovus Cancer Care it was possible to speak to many potential users and perform 2 user experience studies. These were needed to begin to understand user requirements and to gauge current levels of awareness of the terminology and functionality of the tool.

There are numerous models which can be used in user centred design, and they all have their benefits (Rogers et al, 2007, ISO 13407). By using these models, it was possible to keep the user at the centre of the design and ensure that the final outcome was one which could be utilised by the intended user. There has been much research on the effects of different methods which can be used to investigate usability, such as questionnaires or focus groups. The literature review explores many of these methods and discusses their appropriateness in a range of application areas. Initially several meetings took place where user requirements were specified, and the scope of the project was discussed. There are many people within Tenovus Cancer Care that could utilise this tool, and establishing what tasks would be completed and who the users may be was critical to the project. Following the initial meetings, user experience study one (UX1) took place which consisted of a questionnaire, one-to-one sessions and a focus group. The purpose of this study was to show different types of users the initial iteration of the tool and to gain as much feedback as possible. By using different evaluation methods, it was possible to gain a good understanding of the 'typical' user (job title, IT literacy, etc.) and gain valuable feedback on the initial design.

The results of UX1 were collected and analysed, this allowed for a redesign of the 2SFCA tool, and a second UX study (UX2) was completed using similar methods. The results of the UX2 study allowed for final changes to be made to the tool and testing could be completed. The results highlighted several problems with the software and showed the difference between the language used in informatics and geography, and those terms familiar to non-experts. The results also highlighted the learning curve that the user experienced and confirmed that they were happy to learn new tools as long as they could understand the reasons for it.

Finally, a discussion of the process and some of the key findings are highlighted. These takeaways allowed the project to focus on delivering a tool which can be used rather than focus on functionality or performance, which seems to be the case with most of the geographical FOSS tools investigated in section 4.2.1. Using these methods has the potential to increase the use of the tool, and encourage the use of the software solution on measuring access to healthcare services in a range of scenarios.

5.2 Literature review

5.2.1 Introduction

Usability studies are a mix of design, computing and social science and they use aspects from each of these specialisms to design and evaluate computer systems. This is an area of computing which is often overlooked but it can be the difference between project success and failure. There are multiple different approaches to usability studies and it would be impracticable to discuss each one. Instead, several models are discussed as well as many of the more popular methods for data gathering and usability evaluation. As with many aspects of this project this is not a scenario where one method is the correct one and the others have no value, but that many different solutions could have been used and several methods had to be selected which would follow the selected model and provide meaningful results.

User centred design has been developing since the 1980s when two books were released (Norman and Draper, 1986; Norman, 1988). These stressed the importance of the user in the design process and offered some suggestions on how to design systems with the user at the centre of the design, making systems easy to use and understand. The development of the field continued, and with time many areas of specialism have been studied. Many of the models used today can be traced back to work from the late 1980s and 1990s.

For this project to be a success it is important to have a thorough understanding of the user requirements and how the user interacts with the tool. This is central to any good design; Garrett, (2011; p. 10) suggests that “the more complex a problem is, the more difficult it becomes to identify exactly how to deliver a successful experience to the user”. This section explores the history of usability studies and how these have progressed over time.

Many models have been designed to keep the user at the centre of the design process, and some of these are explained and evaluated. There are many ways in which it is possible to conduct user experience studies and some of the key elements are discussed in more detail. Considering the scope of this study it is important to use methods which maximise the results with minimum impact to the participant.

5.2.2 UX models

There are a number of models that have been presented as a way to keep the user at the centre of the design process. There are common themes through many of the models and the study by Nassar, (2012) highlighted that there are 7 commonly used criteria, which may have different names in separate models but are used in most of the models, these are:

- Consistency
- User control
- Ease of learning
- Flexibility
- Errors management
- Reduction of excess
- Visibility of system status

The International Organisation for Standardization (ISO) brings together experts in many different fields to agree standards across different industries; and within human computer interaction they have published several models to follow to ensure good quality standards are met within the industry. Figure 5-1 shows the ISO 13407 flow chart which suggests how usability can be central to the software design process. The iterative nature of the model is key and suggests showing the project to potential users as often as is practicable to ensure a sound design which is suitable for the intended user and limit any assumptions.

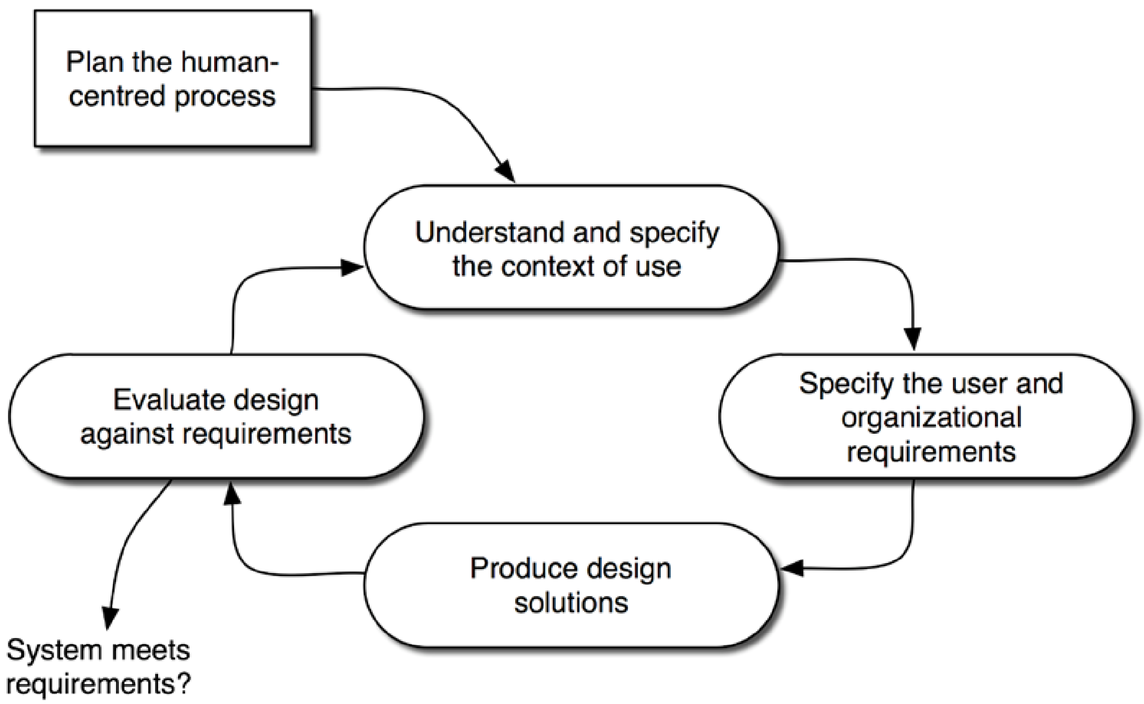


Figure 5-1 ISO 13407 flow chart of the iterative design process

Another framework created for user evaluation is called 'DECIDE' (Preece, Rogers and Sharp, 2004) and the key points can be highlighted with six steps to be used iteratively:

- Determine the goals
- Explore the questions
- Choose the evaluation method
- Identify the practical issues
- Decide how to deal with the ethical issues
- Evaluate, analyse, interpret and present the data

ISO 9241-210 is a framework with six key principles to follow:

- Iterative design process
- Involve users in the entire design process
- The design is driven through user-centred evaluation
- A multidisciplinary team
- An explicit understanding of the user and the task
- The design addresses the whole user experience

Figure 5-2, the ISO 9241-210 flow chart shows how these values should be implemented into the design, as with ISO 13407 this method is iterative and suggests going through the steps as often as needed until the user is comfortable with the design and it is fit for purpose:

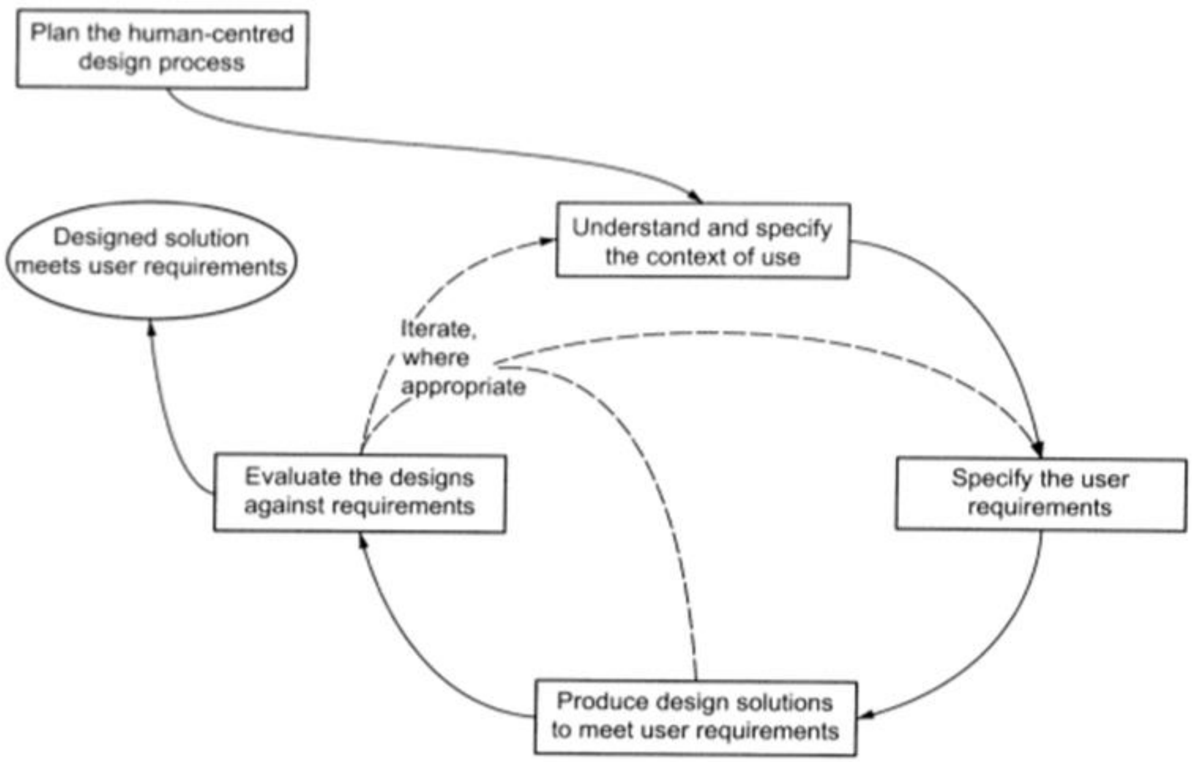


Figure 5-2 ISO 9241-210 flow chart of the iterative design process

5.2.3 Usability studies in FOSS

How usability is used within FOSS development has been studied from several different angles. The research suggests that the use of usability studies in FOSS development is not seen as it is in the proprietary software companies. Kay, (2010) consulted with 27 people working on FOSS projects and found that although most had a good text book understanding of UX techniques the study recommended a model be created to assist FOSS developers get more from usability. Elhag and Abushama, (2009) conducted research concerned with migrating to a FOSS environment and noted that UX is not commonly used in FOSS development. This can cause an issue when people are deciding to move from a proprietary system. Benson, Muller-Prove and Mzourek, (2004) investigated the use of UX on large FOSS projects such as NetBeans and OpenOffice.org and suggested that the way open source projects normally work makes usability studies difficult, but suggested that corporations using these tools have a role to play in assisting with usability studies to help assist the end user.

5.2.4 Discount usability

Although the user is key to usability testing, this is not always the most economical way to test systems.

Nielsen (1989) discussed discount usability, in the paper he suggested that there are some simple tests which can be completed and although they may not give the same results as a full user test, they highlight 'most' of the mistakes in the design. The paper highlighted three main areas':

- Simplified user testing (focussing on small numbers of participants and qualitative methods)
- Paper prototypes
- Heuristic evaluation

A study completed by Nielsen and Landauer, (1993) evaluated a number of different studies to find the best cost-benefit return for usability studies and concluded that in a medium sized project up to 16 participants were worth their cost and 4 participants provided the best cost-benefit ratio. Using as much resource as possible to assist with a user study is important but for many projects it is impossible to set aside as much resource as is suggested in the models. This type of study shows that even if a small amount of resource is available it is still possible to conduct a user study and get meaningful results.

5.2.5 Usability techniques

Data gathering is a key part of any usability study and there are many ways in which this can be done. The important part of data collection is to triangulate the sources and not rely on a singular source of information. By doing this it is possible to mitigate any risks that one data source is poor or that the data is skewed. As with most studies the results depend on the data and gathering the correct data from multiple sources is key to the project.

Preece, Rogers and Sharp, (2004) say that there are several sources of triangulation such as data drawn from different sources at different times and the use of several models to ensure that all the points are being completed, but the most common form of triangulation is that of different methodologies. They suggest that to get a true triangulation is very difficult but conclude that it is good practice to use more than one technique to collect data.

Focus groups are led by a facilitator and allow people to discuss issues in a social setting. The facilitator will try and keep the conversation on topic whilst gaining knowledge from the conversation. Nielsen, (1997) suggested that although the information from focus groups is valuable, it should only be used in conjunction with other data gathering processes. Preece, Rogers and Sharp, (2004) suggested that focus groups allow diverse issues to be discussed which may be missed in other formats, and that people form opinions in social settings.

Questionnaires have been used as a data gathering method in many different situations from market research to the social sciences for many years, and there have been many different studies into how a questionnaire should be constructed to gain the most usable results. Oppenheim (1992) discussed many of the options in questionnaire design. The Likert scale is used to measure opinion and is commonly used; it is a statement with a list of options where the respondent picks the most suitable for their experience. The use of an online questionnaire has been shown to gather more open results from participants (Chang 2004). Giving the participants time to reflect on their own and make anonymous responses was important to ensure that the user had a platform to convey any feelings they may have been uncomfortable discussing face to face. There are also benefits attached to response times as Sue & Ritter (2007) suggested as well as the ability to easily automate the answers from participants. Preece, Rogers and Sharp,

(2004) suggested that gathering some personal details such as age, gender or other attributes will help give context to the answers in the questionnaires. This would assist, but they suggest that only contextually relevant data should be collected.

Interviews and observations have been used in many different settings from journalism to psychology for many years, and there is a great deal of literature on how these should take place and what should be asked within them. Leech, (2002) suggested that prompts are key to a semi-structured interview and can assist in eliciting the required response from the participant. Hove and Anda, (2005) interviewed 12 software engineers multiple times, they suggested that interviews are expensive and that it may be difficult to establish if the interviewee viewed the interview in a positive light.

Evaluation can be done in many ways, from statistical analysis to response times on a computer. As with many strategies with usability there are many different ways that this can be done. Ledo et al., (2018) examined 68 papers and highlighted some common techniques; demonstration, usage, technical performance and heuristics. Preece, Rogers and Sharp, (2004) suggested that the following points are required when evaluating results:

Reliability – How well the method used provides the same results on different occasions.
Validity – Does the method evaluate what it is meant to.
Ecological Validity – How the environment effects the evaluation positively or negatively.
Biases – Does the interviewer consistently miss a type of behaviour.
Scope – How far the results of the study can be generalised.

5.2.6 Summary

It is difficult to cover all aspects of usability in this section, but by focusing on the major methods and models it is possible to conduct a valid user study which informs the design of the tool created. The literature shows that although there are many different ways a study can be done many of the models have consistent themes, and using a model benefits the end design. Using questionnaires, focus groups and one-to-one sessions it is possible to triangulate and help remove any negative findings that may have occurred by only using one technique, and whilst there are many more techniques that could be used the scale of the study needed to be carefully considered to ensure time and resource were used efficiently.

5.3 Tenovus requirements

A meeting was held with the key stakeholders (Head of Research, Head of IT, Data controller and representatives from the mobile unit department) in the project to understand the requirements of Tenovus and discuss how a tool such as this would be used within the business. The main requirements highlighted were:

- Be usable by non-GIS experts.
- Run within their firewall to mitigate any security risk.
- Be able to filter the data so that they can target specific metrics.
- Be flexible and assist with other areas of their business (Shops, Financial support, Choirs).
- Use different metrics (age, gender) to focus on different parts of the population.
- Be able to use the tool for planning and site selection.
- Use the tool for unit optimisation.
- Create visualisations which can be used with their supporters and during interactions with external stakeholders (health boards, supporters).

This functionality is new to researchers in Tenovus and so it is important to provide a wider understanding of the importance of these solutions. In order to examine the potential of these software tools several scenarios were discussed to help them visualise the advantages of such an approach to address their business needs.

5.4 Super user

This type of analysis is inherently complex, and there are certain tasks that require someone with some computing knowledge to complete. For this tool to be usable by non-experts it requires that several tasks are completed prior to the user having access. These tasks include setting up a spatial database and introducing additional data to the database when required. Some data such as census data and road network data need to be updated periodically, and having a super user allows these tasks to be completed prior to use.

It is important to acknowledge that the documentation (appendix C.2) for a super user will need to be different and much more detailed than that for a typical user. It needs to show the user how to download the data and insert it in the correct format into the database. There are other tasks that the super user is required to complete in order to set up the program initially, and all of these steps are discussed and documented thoroughly.

Having the tool set up in this way enables many people to access it, as opposed to a single user doing all of the analysis, and this has benefits to many organisations. The super user also has the ability to amend the code to ensure the tool stays current, and has the option of streamlining their requirements based on the current version of the implemented solution.

5.5 User Experience studies

5.5.1 User Experience study one

The user experience study took place in June 2018 at the Tenovus Cancer Care offices in Cardiff. Eleven semi-structured one-to-one sessions, a focus group and several meetings with key stakeholders took place. The aim of the study was to understand who would be using the tool and what their capabilities are. To achieve this in the one-to-one sessions the participant completed a task using the tool, and an online questionnaire regarding their experience, and a focus group took place where the participants were encouraged to discuss their experience and provide feedback.

The questionnaire captured some general information. It was a platform where participants could comment with no external influence as it was an online survey completed when the participant was alone at their desk. An online questionnaire developed on www.surveyplanet.com was used to ask the questions. In general, a Likert scale was used as this is easier to design than some others; it is often used to measure attitudes (Oppenheim, 1992).

Semi-structured one-to-ones consisted of the participant completing several documented tasks with the assistance of a helper, who could assist with any problems and make notes about any issues that were found. The one-to-one allowed for the participant to gain a better understanding of the project, and any questions that they had were answered and discussed. The sessions served to explain to the potential user more about how they could utilise spatial data and to see if they had any suitable suggestions of how a tool like this may be of assistance in their day-to-day role. Finally, a focus group was conducted with the participants. This was a good platform for discussion; it allowed the participants to talk about the project together and voice any concerns they had and gave them space to talk around the subject whilst encouraging each other to feed in suggestions.

5.5.1.1 One-to-ones

A total of 11 one-to-one sessions took place with one cancellation. Each session lasted between 30 minutes and an hour and it was possible to meet with all levels of staff from most departments within Tenovus. The sessions consisted of an introduction discussing the project and trying to relate it to their work. The participant was then asked to follow a set of instructions (appendix D.1) which requested they input the data into the tool. Each section was explained whilst they were completing it. They were using the mobile unit data as supply, with LSOA total population as demand and a 30km catchment area. Once the tool had completed the calculation the participant was asked to use QGIS to visualise the results, a join was completed and the results were mapped and discussed. Once this was completed the participant was asked to fill out an online questionnaire at their desk.

The main points observed during the one-to-ones included:

- The majority of staff did not know what the server was and that this was confusing for them to complete.
- More than half of the participants made spelling mistakes, and they would all prefer the tool to either auto-complete or have dropdown boxes.
- Help buttons next to each input explaining in detail what is required would be very useful.
- The labels for the data and the sections should be as simple as possible.
- The documentation needs more pictures and needs to be very simple.
- QGIS would be useful for other parts of their role.
- The QGIS section was complex.
- The display needs to be bigger, font and size of boxes.
- The majority of the staff would not be able to use the tool on their own and some guidance would be needed.

These sessions were valuable to enable a good understanding of the IT literacy of the Tenovus staff, and for different individuals to know more about what the project is and how it may be of assistance to their team or their particular role. The above highlighted some really important issues that would not have been considered, and these were addressed in the second iteration of the tool (see section 5.6.2).

5.5.1.2 Questionnaire

Surveyplanet.com was used to create an online questionnaire. 14 questions were asked giving a good understanding of how the tool is viewed, and the general IT literacy of the Tenovus staff. The full list of answers can be viewed in appendix D.4, the answers from the questionnaires were used to guide the development of the tool and ensure it is suitable for the intended user. Completing an online questionnaire allowed, the participant to provide answers with no pressure or external influence and has been shown to provide more insightful results (Chang, 2004).

The first 3 questions asked some personal information to establish the job title, age and IT literacy of the participants. The positions of the participants within Tenovus varied greatly with PA, Head of Care, Marketing Manager, Research Officer, Prevention Lead, Prevention Campaign Advisor, Data Analyst, Choir Leader, Research Engagement Officer and Office Administrator all completing the study which highlights the different type of user that could potentially be asked to use the tool to perform the analysis. Question 2 established the age of the participants and found that there were four participants from 22-30 years old, four from 31-40 years old and two between 51-60 years old which again helps to show a varied potential user base. Figure 5-3 shows the questionnaire response relating to perceived IT literacy; 50% of participants felt they were a medium level of literacy and used IT for work and were comfortable with the different Windows programs (Excel, Word and PowerPoint). Several users identified themselves as being comfortable using more complex programs such as databases and one user reported being able to program or manage databases.

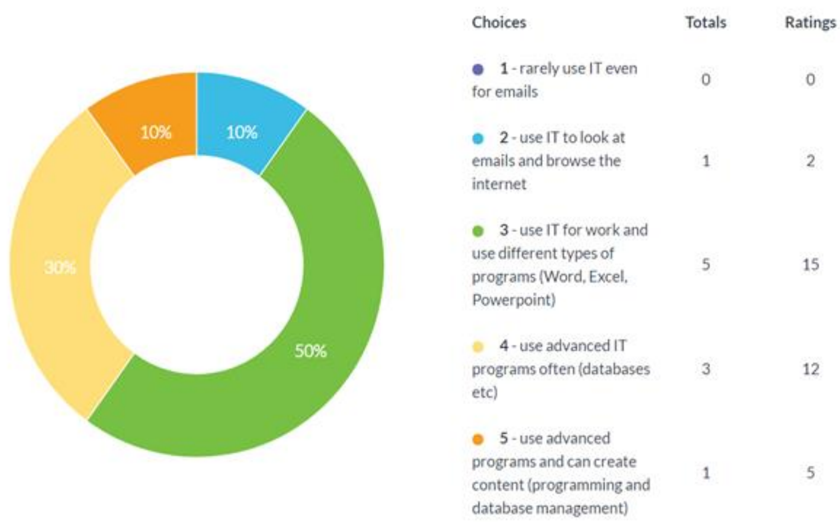


Figure 5-3 Question 3 response

Question 4 asked 'How did you find completing the task?' and using a 5 point response 70% found it to be moderately difficult. This is understandable when compared with the notes taken in the one-to-ones and within the focus group. Figure 5-4 shows the breakdown of the responses.

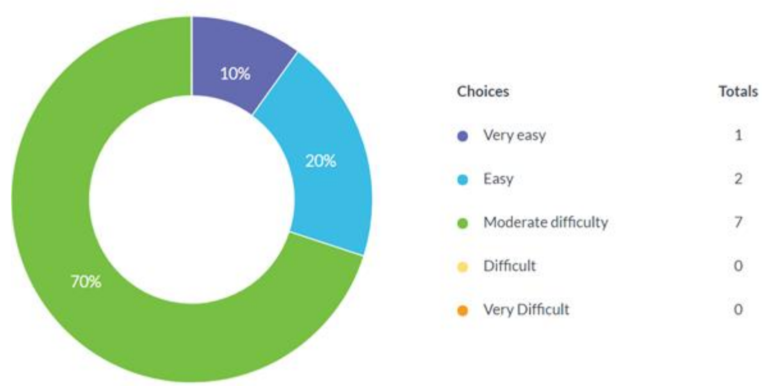


Figure 5-4 Question 4 response

Question five asked 'Did the tool feel familiar to other programs that you use?' and 70% suggested that the tool felt unfamiliar, the remaining participants said it felt familiar but none responded that it felt very familiar. The aim of this was to make the user comfortable, and creating a tool which feels familiar to them and where they know how to find help will be key to the success of the project.

Question 6 asked if the user found the tool intuitive, this question split the participants and 50% said it was not intuitive, 40% thought it was intuitive and 10% believed it was very intuitive. 50% is a large number of potential users who believe the tool is not easy to understand, and this could impact the uptake of the tool within the organisation. Identifying that it needed to be more intuitive allowed for further development of the tool to correct this.

There were several stages to go through in the user study as can be seen in appendix D.2 and mapping played a large part so that the participant could see the result. The next question asked 'Did you find creating the map difficult?'. Figure 5-5 shows the results and only 30% of participants found it easy to complete. This needed to be a higher number for the project to be a success and to aid with uptake.

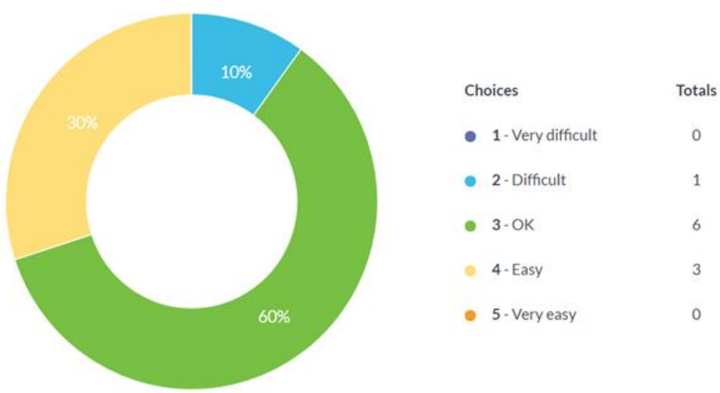


Figure 5-5 Question 7 response

In question 8 the participant was asked if they could complete the task again without assistance and this led to 50% saying that they could complete some of the task and only two respondents feeling as if they could complete the task on their own. One participant responded that they would be unable to complete the task on their own and the remaining participants said they could complete all or most of the task.

Question 9 asked if the participant ever used database systems and 50% responded that they often use database systems with the remaining saying they use them rarely or never. There are two in house database systems in use at Tenovus which were used by many of the staff in most aspects of the charity's business. Question 10 asked if the participant could see any benefits to this tool being available online and 50% were unsure, 30% yes and 20% said no.

Question 11 asked 'Would the additional functionality of QGIS be beneficial to you (creating maps and visualising spatial data)?' Figure 5-6 shows the responses to this question where 80% of respondents believed that QGIS would be beneficial in other aspects of their work. This was an important question as it helped lead the project and assist in how much functionality the user could access within a GIS.

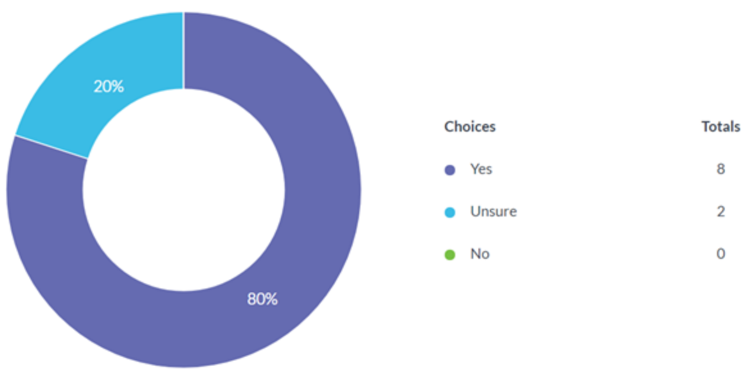


Figure 5-6 Question 11 response

Working in computing or geographical environments may lead to the use of a specific vernacular with many acronyms used by most in these environments. With this in mind it was important to ascertain if the language used in the tool was easy to understand by the participants who may have less exposure to this type of language. Figure 5-7 shows the responses to the question and highlights the fact that half of participants did not understand any or little of the language used and none of the participants understood all of it. This highlighted the need to make the language accessible to non-experts in the field. Without discussing this with the participants it would have been difficult for the project to understand what language is commonly used in non-GIS settings.

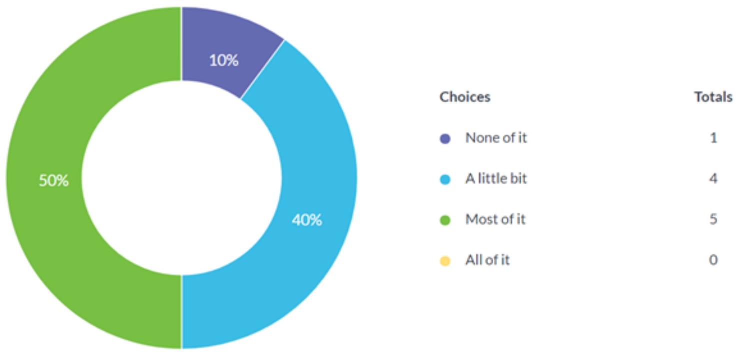


Figure 5-7 Question 12 response

Question 13 asked 'Was the tool nice to look at?' and while 60% found it to be nice or acceptable a large percentage of the responses suggested that it could be better. Question 14 tried to establish if the tool was nice to use and 80% said the tool was very nice, nice or acceptable there were 20% who believed the tool could be better to use.

These questions were designed to be replicated at the second user testing session to see if there was measured improvement over time. This gave an idea of the typical user and could be used to ensure that the tool is pitched at the correct audience.

5.5.1.3 Focus group

The focus group consisted of 4 participants who all took part in the one-to-one sessions, and it allowed them to discuss ideas between them. Most of the conversation confirmed what had been said before in terms of creating good documentation and making the tool as easy to use as possible. It was suggested that videos to demonstrate the tool would be useful as well as helping them to understand what the tool can do for them. It would be good to create some maps for each team and talk them through the possibilities before the project ends. The scheduling of the focus group was poor as it was held at 15:00 on a Friday and there was a skeleton crew in the office.

The group discussed if categorising the data into the different teams would be beneficial to them so that they would know which data to look at. QGIS and its other merits were discussed in relation to each individual, and some of the different tasks that it can complete were mentioned.

5.5.1.4 Summary

In general, the feedback was positive and everybody was happy to take part in the user study. The main points to be considered moving forward are:

- Good documentation which shows users step by step what they need to do.
- Larger size and font so that it is readable.
- Remove the possibility of spelling errors.
- Make the QGIS part of the process more straightforward (remove the join) and consider making a less busy interface.
- Increase the visibility of the project within Tenovus.
- Hide the server connection details.
- Add help sections or prompts so that the user knows what data goes where.

The next iteration of the tool took these points and created an improved user interface (UI) which could be further tested on users within Tenovus.

5.5.2 User Experience study two

The user experience study took place in September 2018 at the Tenovus Cancer Care offices in Cardiff. Eleven one-to-one sessions, a focus group and meetings with the key stakeholders took place. The aim of the study was to understand who would be using the tool and what their capabilities are. To achieve this the one-to-one sessions asked the user to complete a task, and then complete an online questionnaire about the task and some general questions. A focus group took place where the participants were encouraged to discuss their experience and provide feedback.

After carefully redesigning the tool to address the issues highlighted by UX study 1, a second iteration of the tool was created. This tool was then used to perform a second user study. UX study 2 used a similar format to UX study 1 except for during the one-to-ones much less assistance was provided, and the participants were mainly monitored to see how they could follow the documentation and interact with the tool.

The results of UX study 2 were positive and showed that this iteration of the tool was much easier to follow for the users; after 2 attempts most participants understood the process. The study did highlight that users didn't fully understand the scope of the tool, and that this needed to be addressed in the documentation. They also didn't understand what 2SFCA is and would have liked to get a better understanding before they fully utilise the tool. The documentation also needed to be easier to understand.

5.5.2.1 One to ones

A total of 11 one-to-one sessions took place, with one cancelation. Each session lasted between 30 minutes and an hour and it was possible to meet with all levels of staff from most departments within Tenovus. The sessions composed of an introduction, discussing the project and trying to relate it to their work. The participant was then asked to follow a set of instructions which requested that they input the data for the tool; each section was explained whilst they were completing it. They were using the mobile unit and choir data as supply with LSOA total population and cancer incidence as demand, and 30km/40km catchment areas. Once the tool had completed the calculation the participant was asked to use QGIS to visualise the results. Once this was completed the participant was asked to complete an online questionnaire at their desk.

The main points observed during the one-to-ones included:

- The documentation is much improved but needs to be split into sections, with big titles and subtitles to clarify.
- The tool was much easier for the participants to follow and by the second task they were able to get through it very quickly with some not needing the instructions.
- Most people were unable to envisage a scenario where they could use it. This needs to be addressed so that individuals are able to understand the tool's potential.
- Removing the join in QGIS made it much easier for users.
- The data layer names were confusing to the staff. This needs to be addressed to ensure they know what they are looking at.
- Some of the language used should be friendlier, "back" instead of "previous", or have explanations to better explain what is required.

These sessions were valuable to understand how the user experience has progressed and to highlight any remaining issues. The UI was now acceptable to the users and all were able to complete tasks on their own, the focus now needed to be the documentation and helping the user to understand the potential of the tool.

5.5.2.2 Questionnaire

Surveyplanet.com was used to create an online questionnaire, 20 questions were asked and this gives a good understanding of how the tool is viewed and the general IT literacy of the Tenovus staff. The full responses to the questionnaire can be seen in appendix D.6.

The second questionnaire aimed to ask many of the same questions as the initial questionnaire to see if improvements could be measured over the different iterations. Initially the same questions regarding job title (PA, Head of Wellbeing, Head of Research, Marketing Manager, Research Officer, Prevention Campaign Advisor, Data Analyst, Choir Leader, Research Engagement Officer, Office Administrator and Head of Process Improvement) which were similar to those seen in UX study 1. Age was established which showed Four participants from 22-30 years old, four from 31-40 years old one each between 41-50, 51-60 and 60+ years old.

Question 3 asked about the IT literacy of the participants and found that all users have a standard level of IT literacy with several having a high level of computer literacy. 36% used computers for work, 45% used advanced programs such as databases and 2 respondents use computers for database management or programming.

Question 4 asked ‘How did you find completing the task?’, Figure 5-8 shows that the participants found the tool easy to use with 10 saying it was easy and 1 saying it was very easy. This was a big upturn from the initial questionnaire which had a much more mixed response. This response suggested that the changes made to the UI and the documentation assisted in creating a much more usable program.

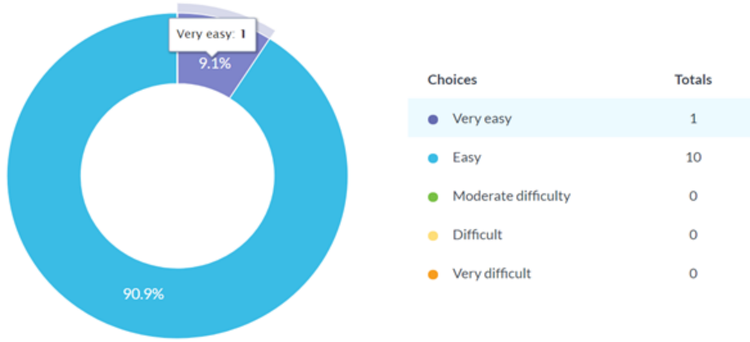


Figure 5-8 Question 4 response

Question 5 asked the users ‘If you found the task difficult please expand on the parts of the process that were difficult or confusing?’ and there was only one response which implied there were problems by a respondent who suggested that on “A couple of occasions I might press the wrong button. If this was the case, I would need guidance as to how to get back to where I was originally to get on the right path.” Although there were not many responses to question 5 one participant suggested that they could get lost at certain points throughout the process. The documentation could be improved to assist the user as well as the UI help prompts.

Question 6 asked if the tool felt familiar to other programs that they may have used? This has shown a large improvement over the results in UX study 1 where 70% of the responses suggested the tool felt unfamiliar to other programs that they have used. This iteration felt much more familiar and almost 65% of participants felt the tool was familiar or very familiar. Figure 5-9 shows the split of the responses and this provides evidence that the tool has improved in usability from the first iteration.

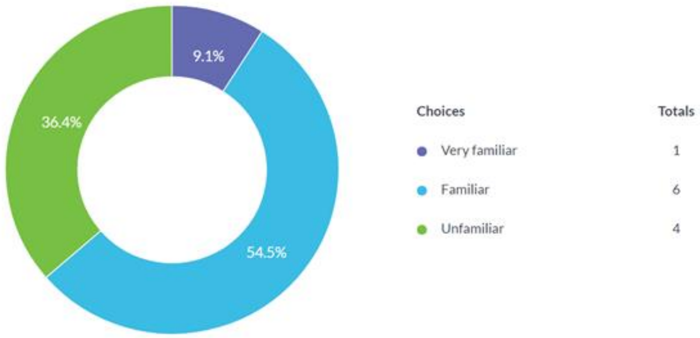


Figure 5-9 Question 6 response

Similarly, to question 6, question 7 highlighted the improvements made to the tool, and asked if the participant found the tool intuitive. In UX study 1 the results were 50% not intuitive, 40% intuitive and 10% very intuitive. The changes attempted to make the tool easy to use and understand, and the results displayed in figure 5-10 show that only 3 of the participants now found the tool not intuitive with a slight increase in those who found it intuitive, and over 27% finding it very intuitive.

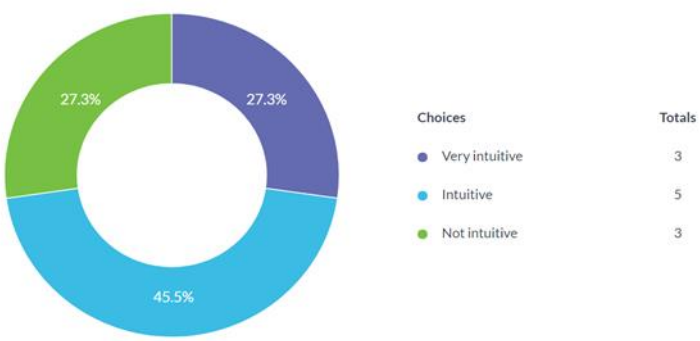


Figure 5-10 Question 7 response

Question 8 asked the participant ‘Did you find creating the map difficult?’ and again the results show an improvement. The changes made to how the tool produces a layer which is easy to add to a map and does not require the user to use joins or advanced functionality in QGIS has been seen as beneficial. 54% found the tool easy to use compared to 30% in the initial study and 45% found it OK. In the initial study the majority of participants found it OK with 10% saying it was difficult.

Question 9 asked the participants to provide additional comments on how they found QGIS:

- Impressive and easy to use.
- Generally good. Might be worth naming the fields, and clearing up the UI just to make it a little more intuitive.
- I found this really straightforward and interesting, I can see how a number of services would use it, the instructions were clear and easy to follow.
- Good program, which I only found slightly more difficult as I had never seen or used it before, but the functionality and user-friendliness of it was clear to see. I don't think it would be long before I would feel comfortable to use and manipulate it in whatever way I needed.
- I found it easy to use with the instructions and once used a few times it should be easy. It would be difficult to get to grips with without the instructions as it isn't like a program I use regularly.
- Useful tool for providing visual examples of geographic analytical data.
- Overall, I think it was excellent. I understood what I was doing, how and why. My only concern would be, as said previously, understanding what I might do to rectify a mistake if I was so reliant on the manual for the pathway to know what to do. A bit more knowledge and a lot more practice should overcome this.
- Once you're familiar with the options you need to use it's fine.

Most users found the QGIS experience was good but noted that some practice was needed, and that the documentation assisted them in completing the tasks. Having these comments away from the focus group allowed for a better understanding of the individual’s experience without feeling embarrassed or cautious of their peers. The positive feedback added weight to the choice of QGIS as a GIS and several comments suggested it could be useful in other aspects of their role.

Following the same path as the initial questionnaire question 10 asked ‘Could you complete the task again without assistance?’ (Figure 5-11) and the responses were an improvement on those in UX study 1. All participants felt they could complete some of it on their own, over 70% could complete most of it and 2 users believed they could complete the entire task on their own. This was a big difference to the first study where 50% could complete some of it and one participant could not have completed any of the task on their own.

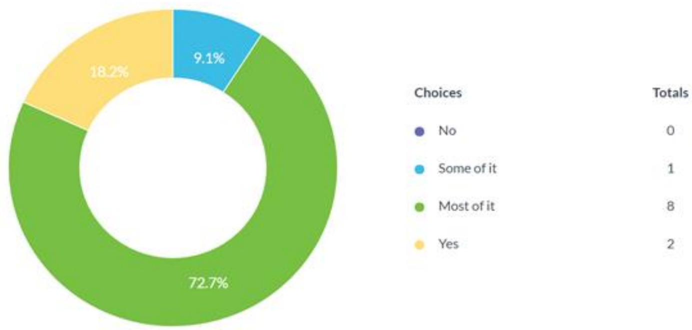


Figure 5-11 Question 10 response

The next two questions asked ‘Do you ever use database systems?’ Over 72% responded that they do often and 27% do rarely, which shows that most users of this tool will have some experience with database systems and ‘Would the additional functionality of QGIS be beneficial to you (creating maps and visualising spatial data)?’ Over 80% of responses said that QGIS would be of benefit to them in other parts of their job.

Question 13 asked ‘Could you run the program again with a different set of inputs?’ and every respondent would be happy to try and run through the process again, with several confident they could perform the calculation and visualise the result. This showed that the tool had been improved to help assist the user in the process and the prompts and help screen would allow the user to feel confident enough to try and use the tool with different data.

Question 14 asked ‘Did you understand the language used?’ and this highlighted an issue, the language was still not easily understood by the participants. Figure 5-12 shows the splits and compared with UX study 1 there is a slight improvement. This issue needs attention and could be addressed by using more simple language or by teaching users the language used in the documentation.

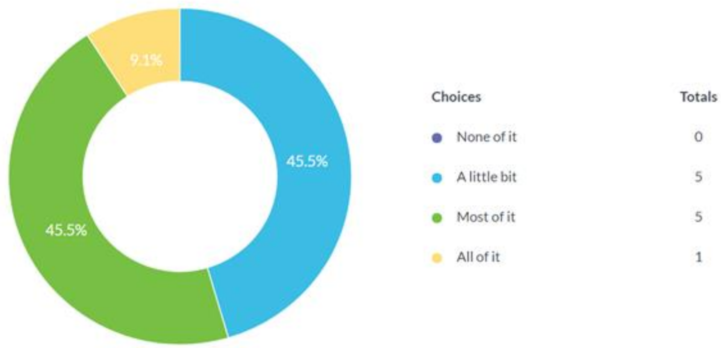


Figure 5-12 Question 14 response

The next question tried to understand how the participants found the tool aesthetically and asked ‘Was the tool nice to look at?’ and this saw a marked improvement over the first iteration. In UX study 1 40% said the tool could look nicer, 30% found it nice and 30% found it acceptable. Figure 5-13 shows that 54% now find it nice to look at and 2 participants found it very nice. There was still room to improve the design, and some more aesthetic tweaks could improve the user experience.

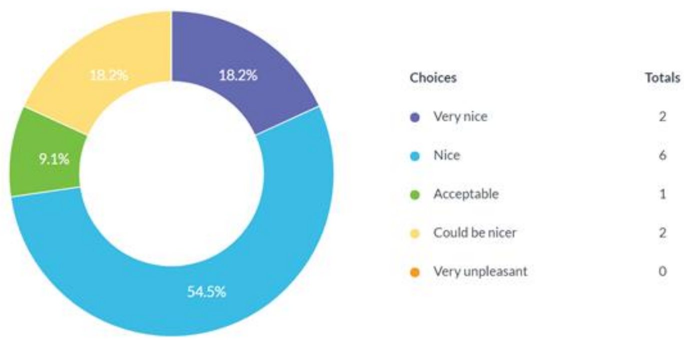


Figure 5-13 Question 15 response

Question 16 asked the user if the tool was nice to use and this has seen a positive shift over UX study 1 with there now being over 70% saying they found it nice which is a change from 30%. 1 user found the tool acceptable compared to 4 users in the previous study, and 2 found it very nice. This was further confirmation that the changes made had had a positive impact on the user experience.

Question 17 asked ‘If the tool was not nice to look at or easy to use please expand on the parts that could be improved’:

- Space out the field names and text boxes. Perhaps could try using Tenovus Colours? Change the field names to make it a little clearer to the end user what each one does. At the moment its very code like, could it be simplified?
- It was quite easy to use and the menu bars were like other programs which I liked. However, it still felt a little bit like the 'back-end' of a program (which whilst easy to use with the instructions) was a little daunting initially.

- The only slightly complicated or that I can imagine would make it harder to use is getting to grips with the naming of the layers. I don't know how easy this would be to change, but this could perhaps make it an even easier tool to use.

Having some direct feedback on specific issues the users found was essential to creating user friendly software. Only 3 participants responded but they gave valuable insight, suggesting that the tool looked “back end” or code like. One user suggested using the Tenovus colour scheme to make it more familiar and the use of better names for layers is something that can be implemented to aid the user experience.

Question 18 asked ‘Have the changes to the tool made it easier for you to use?’ and the results showed that all participants noted an improvement with 63% saying the improvements helped and 36% saying they have helped a lot. This feedback added further backing to the changes and the design direction of the user interface. Question 19 asked ‘Does the tool fit in with the Tenovus brand?’ and over 72% believed it fitted the branding with 27% suggesting it does not really fit the branding. Although most participants found the branding suitable there is still room for improvement.

The final question 19 asked the participant to ‘Please describe a task you might use the tool to complete’:

- Establish the best place to locate a mobile treatment unit or choir.
- To see what the catchment area of my choirs are.
- Mapping our charity shops and trying to select and prioritise the ones to run research engagement events in.
- To see where our sun cream van visits and where we could potentially visit in the future.
- The tool could be used to help us to understand how much ground we cover (literally), how much 'reach' we actually make with each or a combination of services, and potentially where we could go next with a service to increase our coverage. Additionally, a function where we can overlay our services against others that operate in the same area could prove very useful, as we would then be able to understand the different areas' impact and how much of a difference we could actually make. Using all of this could be useful in a number of conversations, from speaking to potential sponsors, reporting and planning strategically - it could be a real valuable tool for us.

- Tracking where patients are in comparison to outreaches for CSA advice.
- Although I like the tool as an extremely useful piece of software, it is unlikely that I would be using this in my current role.
- To analyse whether our services are best placed to provide the best possible impact.
- Three-fold. In the research team I might use the system to map in the future the impact of our research engagement in specific areas. We might try and see the grants we fund in relation to services offered in the area. Finally, as part of internally evaluation, we would use the multiple functions of the tool to be able to visualise the reach of all the service that we offer across Wales.

All but one of the responses could find a use for the tool in their current role. This was with minimal exposure to what the tool is capable of and with further experience it could prove to be much more useful than they were able to establish at this time.

Most of these questions were similar to the original questionnaire and it was possible to measure progress over the two studies. This gave an idea of the typical user and could be used to ensure that the tool was pitched at the correct audience. Being able to measure the progress of the tool in several iterations had allowed the user to be at the centre of the design process and ensure that the tool was fit for purpose.

5.5.2.3 Focus group

The focus group consisted of nine participants who all took part in one-to-one sessions, and it allowed them to discuss ideas between them. Most of the conversations confirmed what had been said before in terms of creating good documentation and making the tool as easy to use as possible. It was suggested that videos to demonstrate the tool would be useful as well as helping them to understand what the tool can do for them. It would be good to create some maps for each team and talk them through the possibilities before the project ends.

The majority of the conversation was around the documentation and how to get people engaged in its capabilities. QGIS and its other merits were discussed in relation to each individual and some of the different tasks that it can complete were mentioned.

5.5.2.4 Summary

In general, the feedback was good and everybody was happy to take part in the user study. The main points to be considered are:

- It was easier for each participant to complete the tasks than the first iteration.
- The documentation needs improvement to make it easier to follow.
- The users need help to understand the scope of the tool and how it can help them.
- 2SFCA needs to be explained.
- Further training and documentation are required for the Tenovus super user.

The UI is acceptable and all were able to use it. The documentation needs work and can be updated before the next visit.

5.6 Results

5.6.1 Iteration 1

Figure 5-14 shows the initial design of the 2SFCA tool created. The initial design focused primarily on functionality and not usability. All of the inputs needed to be completed manually and there was no assistance as to what each input referred to.

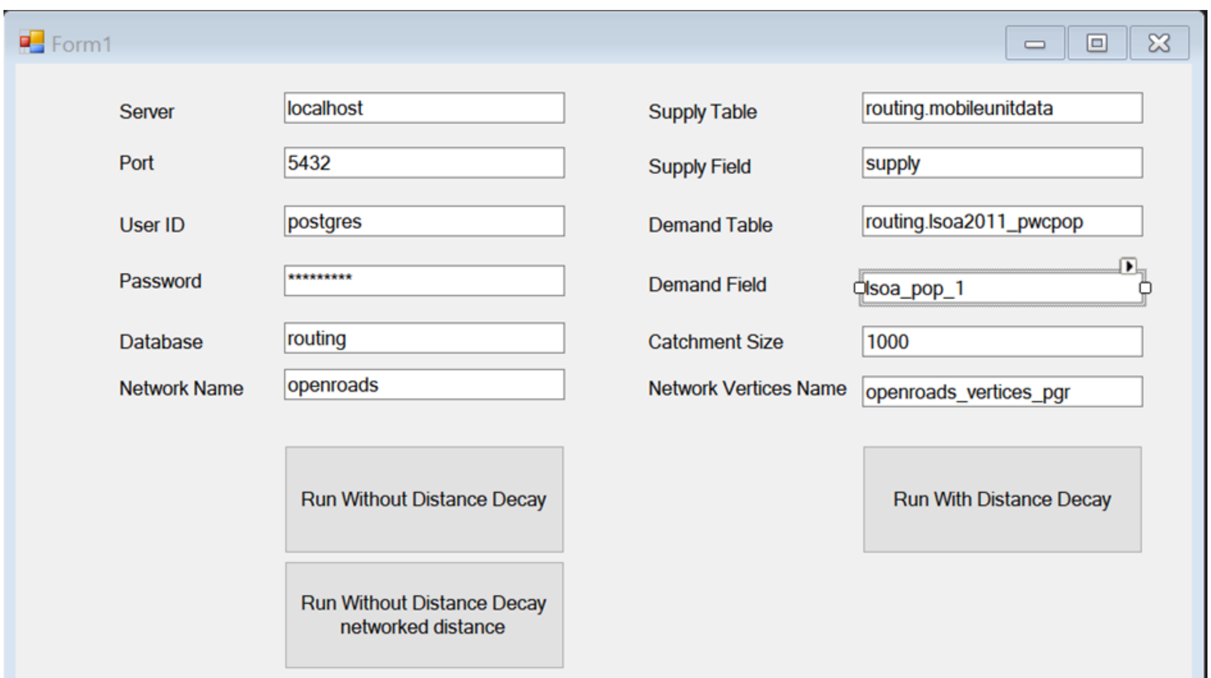


Figure 5-14 First iteration of the tool

The results were displayed in a PostGIS table which required significant editing to display in QGIS and this led to a great deal of confusion amongst the participants. The language used was not correct for the purpose, and the majority of the participants did not know what the different terms were referring to. The documentation (appendix C) did not effectively allow the user to follow the process and it was highlighted that more pictures were required with a more structured process.

In general, the response was that the tool was hard to use and difficult to understand, and more help was required for the user to understand what they were doing and why. Having much clearer documentation was required, and a way to stop the user making mistakes needed to be implemented. On screen help was suggested, and making use of more simple language would assist in the usability of the tool.

5.6.2 Iteration 2

The redesign for the second iteration of the tool (Figure 5-15) encompassed the feedback from the first usability study and tried to focus on the main outcomes which were:

- Good documentation which shows users step by step what they need to do.
- Larger size and font so that it is readable.
- Remove the possibility of spelling errors.
- Make the QGIS part of the process more straightforward (remove the join) and consider making a less busy interface.
- Increase the visibility of the project within Tenovus.
- Hide the server connection details.
- Add help sections or prompts so that the user knows what data is required at different inputs.

The redesign tried to make the tool simpler and easier to understand. This was done by using a larger font and larger buttons where possible. The tool now only required 5 inputs before it could run; the server connection details and road network information were on a separate screen which the users could access if they wished, but if they did not understand it was all set up and ready for use. Using the 4 dropdown boxes the user now chose from preselected data what they wished to run, rather than manually entering the names of each input. The catchment size was limited between 5 and 60km so that the user was guided to a suitable distance input.

The introduction of help on the screen addressed some of the concerns. This allowed the user to read the help section and follow-on screen with no additional documentation. The help button presented the documentation, and hovering over each input prompted the user in what was required.

Previously the user had quite a few processes to complete in QGIS before they could visualise the result; by adding additional process within PostGIS it was possible to make the visualisation more streamlined and simpler.

Figure 5-15 Second iteration of the 2SFCA tool

Figure 5-16 shows the second screen of the tool with the network and server connection details. These needed to be accessible as they may be required by more advanced users, and having the ability to easily manipulate these fields without amending the code could be required. These inputs were required for the tool to run, and by moving them to a separate screen it allowed for the simplification of the main screen for the 'standard user' whilst allowing more flexibility for 'super users'.

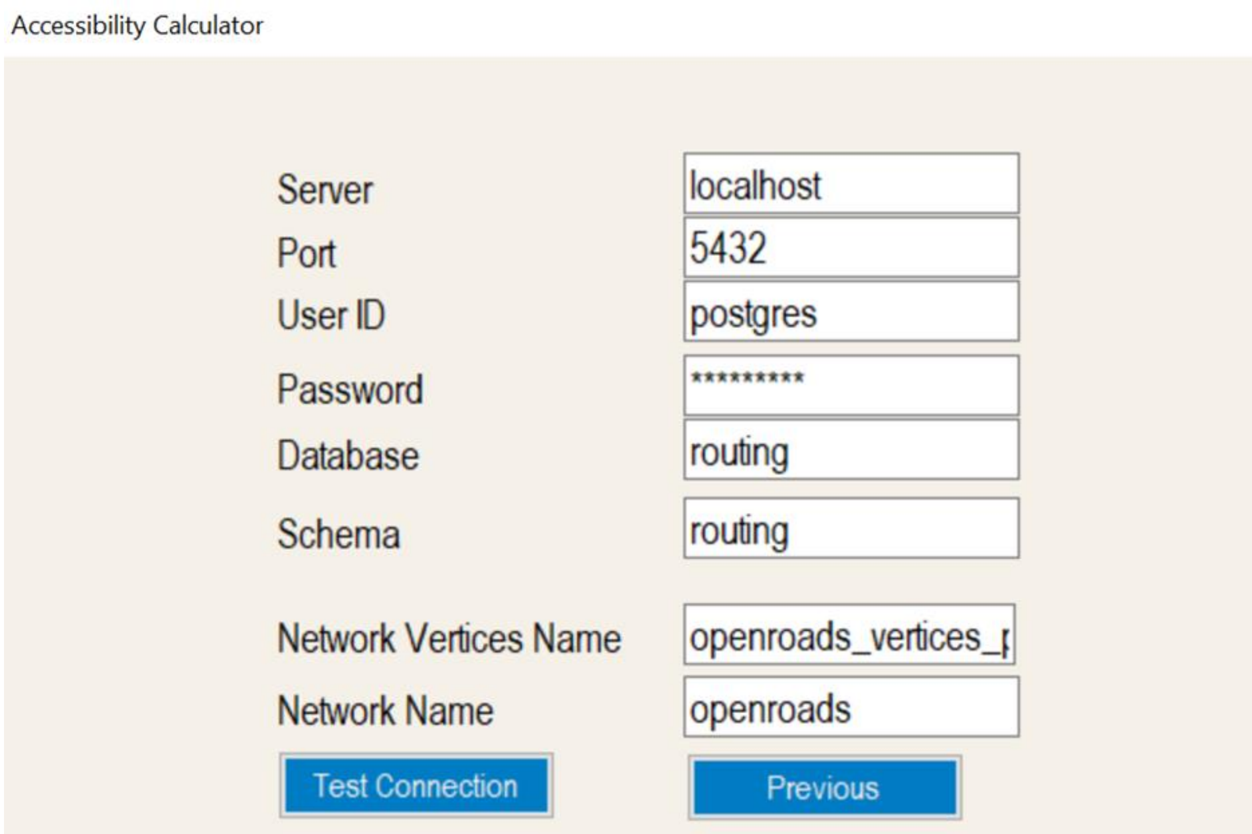


Figure 5-16 Second screen of the second iteration of the 2SFCA tool

The changes made allowed for a much more usable interface and removed a lot of the margin for error. The documentation (appendix C) was greatly improved with better language, more pictures, and highlighted the pictures to allow users to know which stage they had reached. The insights from the user study drastically changed the design of the tool and guided the complexity of the format, language and documentation.

5.6.3 Final iteration

Following the second user study the main feedback points were:

- It was easier for each participant to complete the tasks than the first iteration.
- The documentation needs improvement to make it easier to follow.
- Users need help to understand the scope of the tool and how it can help them.
- 2SFCA needs to be explained, perhaps using a video.
- Further training and documentation are required for the Tenovus super user.

Changes to the tool drastically increased the ability of the participant to complete the tasks, and the feedback was positive. Several changes were requested, mainly in the documentation, to increase the ease with which users could follow it; and also, to explain what 2SFCA is and why it is helpful for the analysis to take place. The final iteration (Figure 5-17) took on points from this and simplified the layout of the design whilst increasing the functionality of the tool. The data analysed from Tenovus showed that the demand side data is Gaussian in nature, so a Gaussian distance decay was included and positioned in the second more complex screen of the tool.

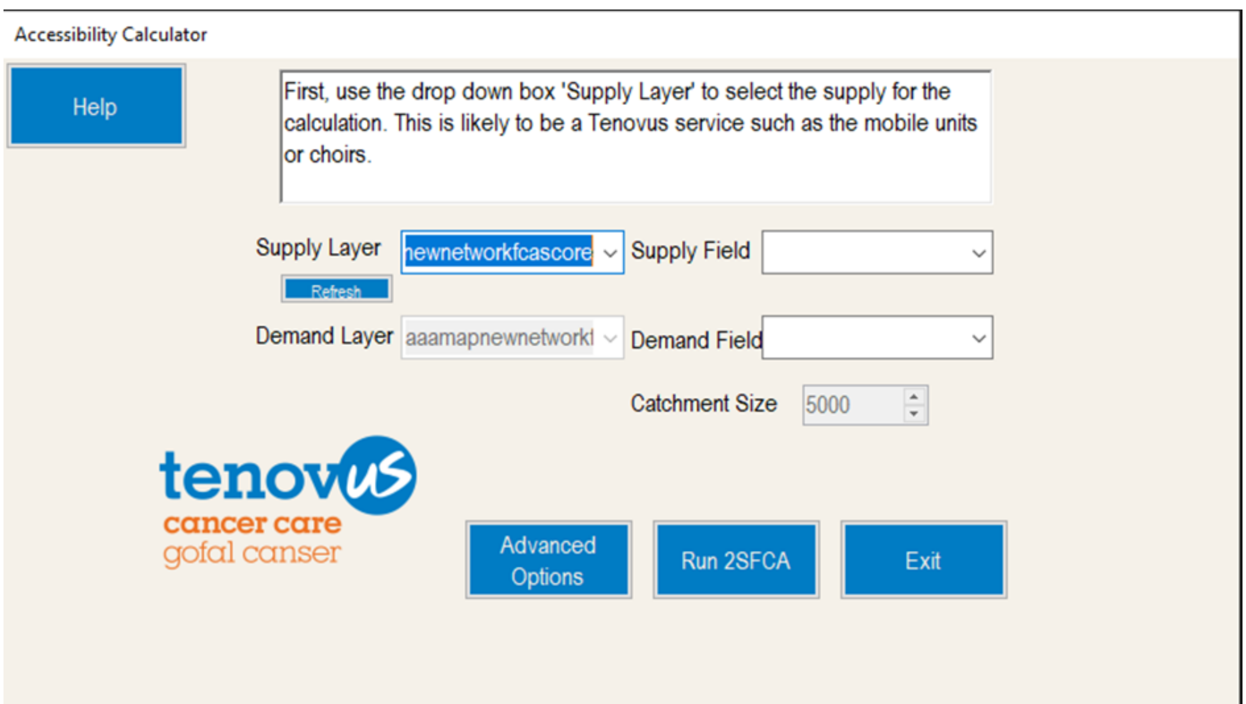


Figure 5-17 Final iteration of the 2SFCA tool

Removing the different functions from the front screen took away several choices for the standard user and was configured to perform networked E2SFCA as standard. This configuration delivered accuracy without additional complexity for the user.

The second screen (Figure 5-18) had a larger amount of detail and allowed the tool to be customised to the needs of the user. This did not need to be completed by a standard user unless they had a specific need for a particular dataset. The increased functionality and simplification of the interface created a more streamlined and powerful tool.

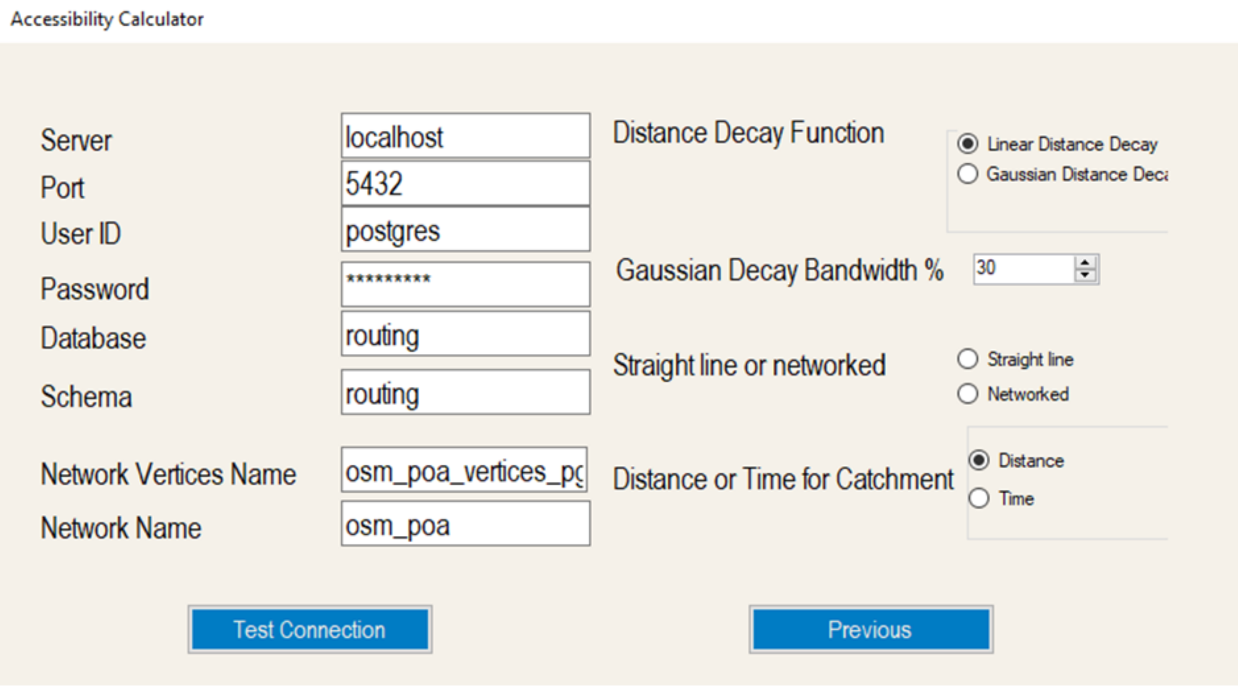


Figure 5-18 Second screen of the final iteration of the 2SFCA tool

5.6.4 Documentation

Documentation is key to the success of this project and ensuring that it is used. The user studies highlighted its importance, and one of the key aims of the project was to create a tool which can be used by non-experts. There are 3 documents which address the non-expert user, the super user, and one which gives 2SFCA some context and assists in interpreting the results.

5.6.5 User documentation

The user documentation (appendix C) uses images and text to assist users through the inputs in the program. Initially the tool is displayed (Figure 5-19) and each function is discussed. Each of the numbers is expanded upon so that the user understands it.

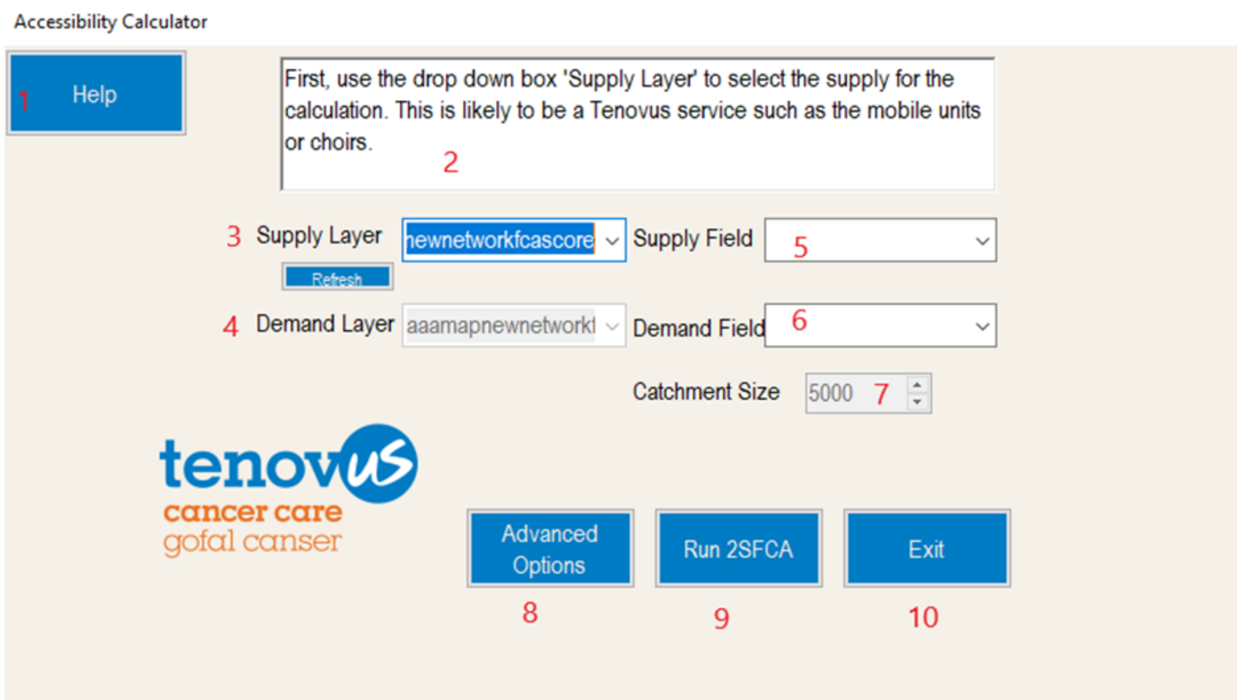


Figure 5-19 Overview of Tool functionality

Following the introduction to the program, the documentation walks the user through each stage of the process. This is done by using a screen shot (example Figure 5-20) which highlights the areas in focus and explains what actions are required

The user is then shown the advanced settings page in less detail. There is a lot of functionality and it could be confusing to go into too much detail at this stage. It is important to show each of the functions and give them context.

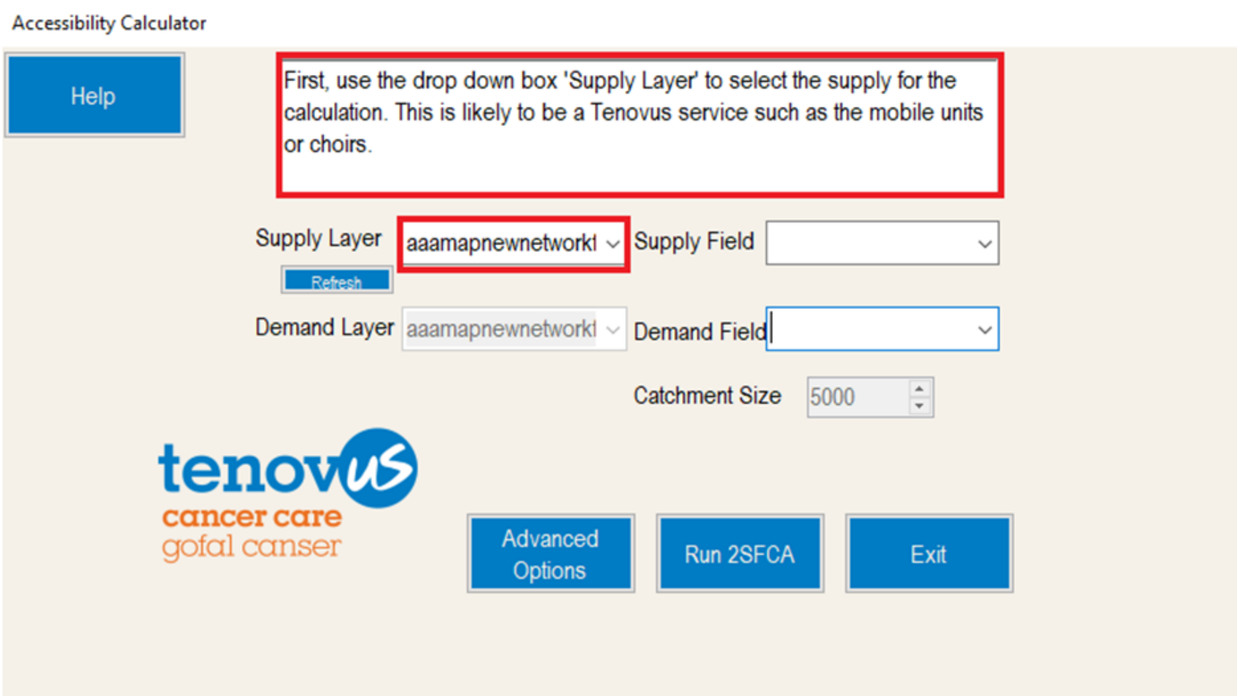


Figure 5-20 Areas in focus and actions required

The language used is simple and conversational to engage the user, jargon is limited and expanded upon where necessary. The user studies highlighted the importance of language and visual cues which have been central to the documentation.

The first thing that needs to be entered is the supply layer. This refers to the table within the database which holds all of the information on a supply of services. In this example chemo_all is selected and this dataset holds all of the data regarding the different sites in Wales which provide chemotherapy services. Most of this data is not used in the front end and is required for the program to run (see below).

5.6.6 Super user manual

For a tool such as this to work someone needs to have a decent knowledge of computer systems. It is not feasible for a tool utilising these methods and datasets to be maintainable without a super user. This was discussed with Tenovus in the initial meetings and the solution provided was designed to fit in with their existing capabilities. The superuser manual describes the processes required for installation and maintenance of the relevant programs and the main datasets. Regardless of who uses the system it needs several key data sets to provide useful results.

Initially the documentation discusses the software required and where to download the versions required for this tool. As the software is FOSS it is possible to get it from many different locations. Some of the software is easier to get from one location and use the provided stack builder so that the tools work correctly when installed.

Downloading and installing software

The software required can be downloaded from OSGEO using https://download.osgeo.org/ there are other available sources. I would highly recommend using the PostgreSQL stack builder to do this as you can get most of the packages needed. The program should work with most of the versions of the software but the recommended versions are:

QGIS – v. 2.14 (for visualisation only)

PostgreSQL – v. 9.5

PostGIS – v. 2.3 (stack builder)

Microsoft Visual Studio (to change the code of the program)

pgRouting – v. 2.3.2 (stack builder)

NPGSQL (stack builder)

osm2po – v. 5.2.43 (for transforming OSM data and can be downloaded from: https://osm2po.de/)

Java runtime environment – (can be downloaded from https://www.java.com/en/download/)

2SFCA tool from download or GitHub

There is a good tutorial on the installation of the items 2,3,5, and 6 here: https://www.bostongis.com/PrinterFriendly.aspx?content_name=postgis_tut01

There are several links to external videos or tutorials which can be of assistance in the installation of specific programs, FOSS systems have differing levels of documentation and some external links can be critical in the installation and de-bugging process. It is important that the user is able to customise the program to their specific needs. This is one of the key principles of FOSS and allows the tool to have more impact in more settings. The tool is really a series of SQL commands which could be implemented in many different scenarios, but it is not feasible to expand greatly in this document, and most of those attempting something like this are familiar with IDEs and different programming languages.

Using OSM data can be complex and it needs to be formatted correctly to be used as a routable network. To do this osm2po has been utilised and the different steps required are explained point by point. There are links to tutorials as this can be complex for those not used to working in the command line. This is a process that does not need to be done often and, as section 6.4.2 showed, there is an accuracy increase by using this data over the OS open data provisions. The tool works with other data and the methods for using OS Open roads is included in the code if the super user prefers this.

The demand data is much the same as the supply data and a table is included for reference. A link to census data is shown as this is a good starting point. There is an assumption of knowledge here as these tasks require some technical skill to complete, but with the included instructions and the links to guides it is possible for someone with quite limited technical knowledge to set up a working E2SFCA accessibility tool.

5.6.7 2SFCA explanation

A brief high-level explanation of accessibility and 2SFCA has been created to help users put the tool into some context. The explanation is not designed to discuss the intricacies of the method but to help them to understand the potential benefits of using this method over the population to provider and proximity methods often used.

Although there are many methods, only two examples have been shown here as they can easily be represented with images. Simple language and diagrams have been used to give a very high-level overview. A similar high-level approach has been used for the 2SFCA explanation. The use of images to explain the process aids in the user’s understanding.

Finally, it is important for the user to be able to understand the maps created at the end of the process and a very brief section on interpretation of the results is included. This imagery is important to show the potential of the tool and help convince the user of the advantages of a 2SFCA approach to examine spatial variations in accessibility to facilities/services.

5.7 Chapter Summary

Completing user studies helped ensure that the user was central to the design of the tool. Keeping the end user involved in the design process from the outset enabled the tool to be more focused and allowed for the 2SFCA methods to be potentially used by more people. The studies highlighted the importance of documentation and removing terms that are commonly used within geography or informatics during the course of a GIS project, but needed to be tailored to the particular task in hand. The differences between the initial design and the final iteration are significant; such changes when implemented have increased the functionality of the software solution whilst making the design usable to a wider audience. By using different methods, it was possible to gain different perspectives from potential users and allow for a more rounded response. The one-to-one section of the study allowed the participants to gain an understanding of the tool and its capabilities which was central to boosting the knowledge of the user.

Documentation has a large role in how the software is utilised and installed, and different formats of documentation tailored to particular types of users have been outlined and included as Appendices to this thesis. User studies highlighted specific needs for documentation to ensure that the aims of this project are met. The quality of FOSS documentation can vary and to have high quality documentation should facilitate wider use of the GIS tools.

Chapter 6 Results

This chapter details the results of the investigations described in the methods section. The inclusion of performance testing, implemented case studies and data comparisons are discussed to address the aims and objectives outlined in Chapter One. During the development of the program, it was important to understand how the inclusion of different data sets and algorithms affected the performance of the tool and many tests were completed on performance under different scenarios. These included using data at different spatial scales to understand aggregation effects on the performance of the tool, and the accuracy of the applications of different 2SFCA algorithms using alternative data sources.

A case study was completed to highlight the capabilities of the software solution. A discussion with researchers within Tenovus enabled the case study to incorporate the types of data sets available to a wider group of users and address a wider range of operational tasks that to date were conducted without the support of GIS technologies. For example, utilising supermarket data sourced from Geolytix, several scenarios were explored to understand the effect of moving a Tenovus mobile unit or adding additional capabilities to the current chemotherapy provisions within Wales. The results show how the Tenovus resource affects the accessibility to a wider range of cancer services within Wales.

The tool was used to create map visualisations for Tenovus regarding its current offerings in chemotherapy and lymphoedema, and how these related to the current static provision provided by other NHS providers. These visualisations have been utilised by Tenovus internally to heighten knowledge within the research team as well as externally to show stakeholders trends in the overall provision of cancer services.

As previously discussed, there were many ways that this software could have been implemented. A large part of this research study involved gaining an understanding of the overall influence of some of these underlying factors on overall access to existing and modelled patterns of provision. With this in mind much work was undertaken to compare and contrast the different network algorithms and data. A comparison between Open Street Map and OSM network data with Google maps gave a better understanding of how each network is composed, and the times which the same Dijkstra algorithm takes to compute different routes on each allowed for the correct choice to be made when selecting a network data set. Work was completed on the different network algorithms, and what effect they had on performance and how the program is assembled.

Following on from network comparisons, different testing was completed on the choice of algorithm; whether it be how the program was constructed or the differing choices for E2SFCA algorithm, and the effect that each of these had on efficiency and accuracy. The aim here was to find the 'best' solution which may not always be the most accurate or the most efficient, but simply a happy medium with the best outcomes overall when data availability, complexity of computation, speed of analysis and quality of results are considered

6.1 Testing

6.1.1 Performance and stress testing

It is important for this tool to work within an acceptable time to ensure the user continues to interact with it and that it maximises its value to potential users. Table 6.1 compares the differences between straight-line and networked 2SFCA calculations. Getting the time or distance for a networked 2SFCA calculation takes a considerable number of resources in comparison to the non-networked calculations and this is represented in the times taken to compute. If the data collation and analysis effort is large, then it may be worth considering using a straight-line calculation in the first instance. However, the routed results give a much more realistic representation such that it may be advantageous to allow a longer time to compute (even a relatively large catchment takes less than a minute to compute).

Catchment size	**Straight-line 2SFCA**	**Networked 2SFCA**
1,000 metres	0:0.13	0:50:74
10,000 metres	0:0.16	0:50:92
30,000 metres	0:0.20	0:54:38
50,000 metres	0:0.24	0:58:92

Table 6.1 Network and straight-line 2SFCA speed comparison

Table 6.2 shows the times taken to compute different E2SFCA calculations using the total population of each LSOA as the demand, and the number of beds or chairs at each location multiplied by the opening hours as supply. This shows that as the data gets larger the time taken to compute goes up dramatically

	2SFCA Networked
5,000 MSOA (410) LSOA	05:33.87
50,000 MSOA LSOA (1909)	07:09.14
5,000 LSOA LSOA	41:30.66
50,000 LSOA LSOA	41:37.65
5,000 OA (10,036) LSOA	1:51:13.82

Table 6.2 Time taken to compute data at different size and scale

The main reason for the time increase is the way the algorithm works; it takes all the records from each input and then assigns and finds the nearest position on the line of the road network. This needs to be done for every record. This can be seen in the difference between the 5km and 50km LSOA times which are very similar as there are a limited amount of LSOA centroids, and this number does not put a great deal of stress on a system, whereas when the OA centroids are added the amount of time to compute the calculations is greatly increased.

6.1.2 Comparison with the ArcGIS tool

There is a similar tool available at present which is an add-on in the ArcGIS environment. This tool is not FOSS and therefore does not cross over with the target audience for this project. It is therefore important to understand the differences between the two and to compare the accuracy of the results against each other. To do this several tests were completed using the same data and modelling parameters in both programs and the output data compared. This section discusses the difference between the two outputs and the way in which both programs work.

A direct performance comparison is very difficult as both programs work in different ways. The ArcGIS tool uses just the ArcGIS platform to complete the entire task; there is a built-in network analyst and spatial database capabilities, with the ability to visualise the results directly in the ArcGIS platform. ArcGIS has some very useful capabilities and it enables the Langford et al tool to run differently. For example, the Langford et al tool requires some pre-processing to join the nodes to the network whereas this is done on the fly within this tool. As this is the case a direct speed comparison is not being included in this project because there are too many variables and differences in the modes of operation to make this a useful activity

A series of comparison tests was completed using both programs, and some basic statistics were taken from both. It would be expected that the sum would equal the number of supply points. Table 6.3 shows the difference between the Langford et al tool and the 2SFCA tool developed for this project to 6 decimal places. The 2SFCA analysis was completed on the Tenovus mobile unit locations with LSOA as demand at 5,000 metres. Figure 6-1 highlights how similar the results are across the data set used, and there are only a couple of points which show very slight changes in results.

	AVG	Max	Min	Sum	Total count
ArcGIS tool	0.046459	1	0.008333	26.90002	579
2SFCA tool	0.046794	1	0.008333	27	577

Table 6.3 Comparison 1 of ArcGIS and the 2SFCA tool results

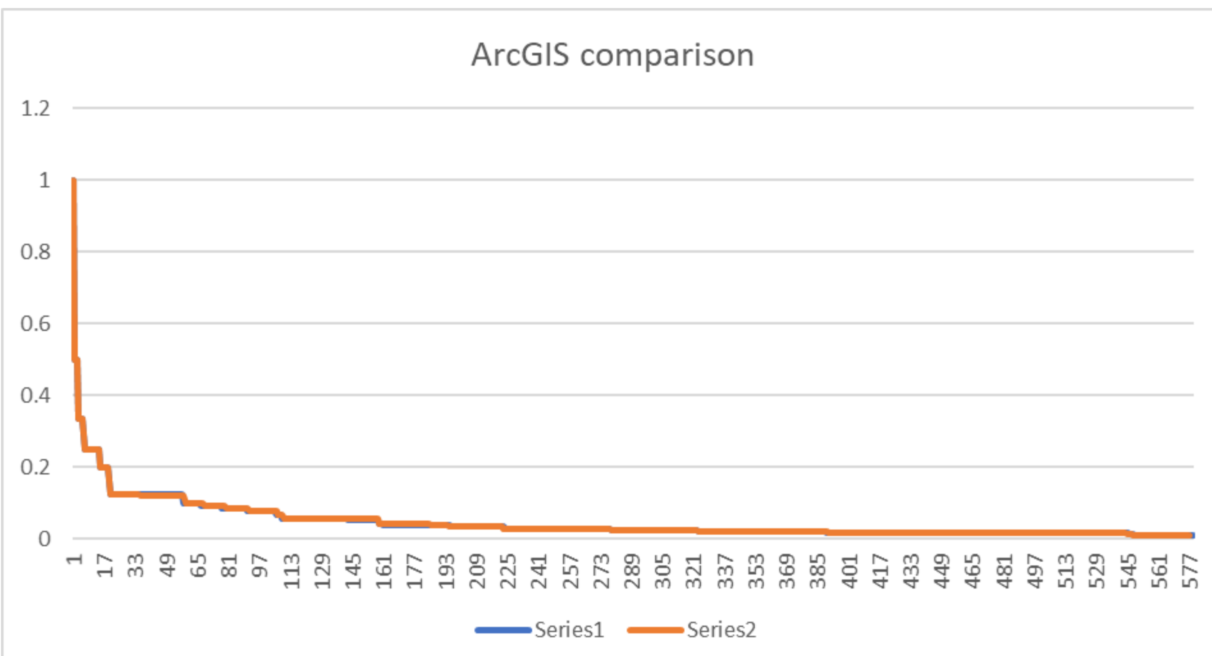


Figure 6-1 Comparison of ArcGIS tool and Tenovus tool where the y axis is the FCA score and the x axis is the series of results

Table 6.4 shows the differences in results between the Langford et al tool and the 2SFCA tool developed for this project to 6 decimal places. For demonstration purposes the 2SFCA analysis was completed on the Tenovus choir locations with LSOA as demand at 1,000 metres. Figure 6-2 shows that at very small scale there are some differences in how the two programs work which would probably be due to how each round different numbers but that the results are similar and similar visualisation patterns are evident.

	AVG	Max	Min	Sum	Total count
ArcGIS tool	0.293103	1	0.111111	16.99999	58
2SFCA tool	0.278689	1	0.111111	17	61

Table 6.4 Comparison 2 of ArcGIS and the 2SFCA tool results

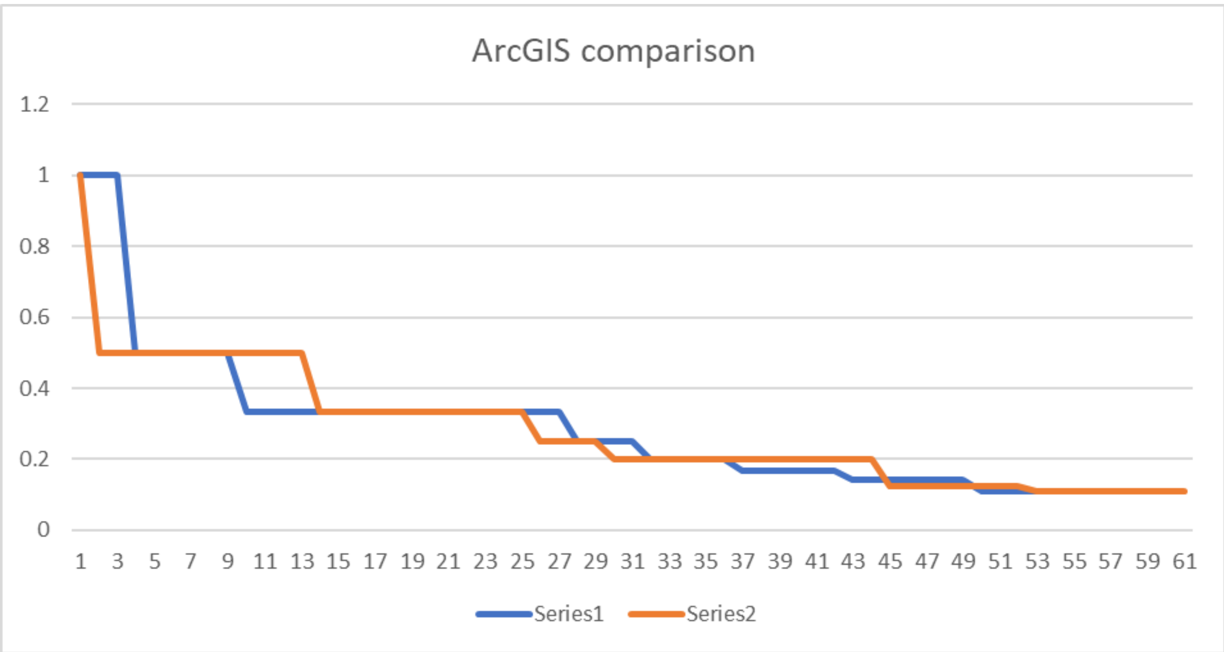


Figure 6-2 Comparison of ArcGIS tool and Tenovus tool where the y axis is the FCA score and the x axis is the series of results

6.2 Case study

For this project to have an impact it needed to be tested in real world situations using real data. After discussions with Tenovus it was decided to run a case study to investigate the use of alternative mobile unit locations, and to understand the implications of adding additional provisions to new locations. A major benefit of being able to run this type of analysis stems from the need to work collaboratively with new partners and demonstrate the effects of implementing additional resources on spatial patterns of accessibility.

The majority of the locations used by Tenovus are large supermarket car parks that have the capacity to offer the 30 spaces the units require to operate. They also have the benefit of allowing a spouse or carer to collect some shopping during the chemotherapy session. A data set of the supermarkets throughout the UK (GEOLYTIX retail points) was utilised and mined to provide a list of large supermarkets in Wales. The criteria for the stores were that they must be over 2,800 square metres with the assumption that any of these megastores would be able to provide adequate parking for a mobile unit. Large retail units have historically offered Tenovus the opportunity to use their facilities free of charge, and they are normally well positioned to access transport links. Such factors may be investigated alongside other considerations which may influence the final choice of sites chosen, such as the ability to get staff for the units which are normally provided by the health board and the logistics of moving units across large distances and how they can be maintained.

The suggested sites were not exhaustive, but they offered an insight into the use of GIS-based accessibility tools to examine existing patterns of accessibility, and how accessibility to either chemotherapy or lymphoedema treatments could be improved across Wales. The aim of this case study was to demonstrate to the research team within Tenovus the potential of the tool to identify new sites to potentially plan services. The locations shown are just an example, and with the knowledge within Tenovus other sites could be identified (sports clubs, cinemas etc.) using these types of techniques.

Geolytix provides a data set of all major supermarkets in the UK. Out of 13,860 shops it was possible to narrow the number down; firstly, by removing the smaller shops which would not have enough parking to accommodate the Tenovus mobile units, and also by removing locations that are not in Wales. Figure 6-3 shows the 79 remaining megastores within Wales that are over 2,800 square metres that could be considered as potential sites for services.

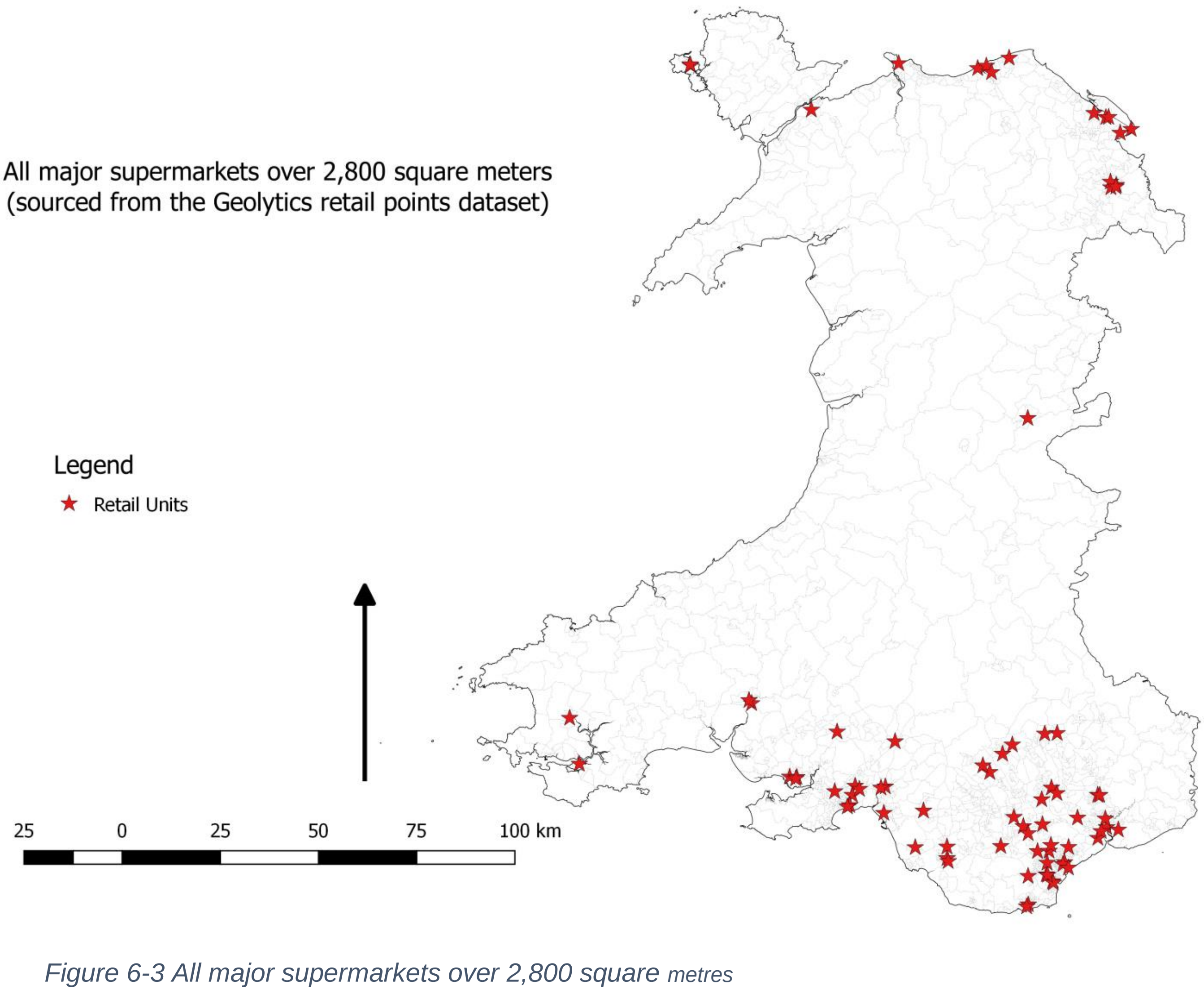


Figure 6-3 All major supermarkets over 2,800 square metres

As might be expected, the majority of locations are around large metropolitan areas in the South and North of the country. This pattern is followed in the provision of chemotherapy throughout Wales as shown in figure 6-4. Having the majority of the services near the majority of the population centres follows logically. One of the main exceptions to this pattern is in Newtown, Powys which is located in the East near the border with England is near the centre of the country between the North and the South and has a megastore and no chemotherapy provision at all. This is also the case in Holyhead. There are pockets of lower accessibility within the large amount of provision in the South East of Wales and Pyle would offer an opportunity to increase accessibility within the area.

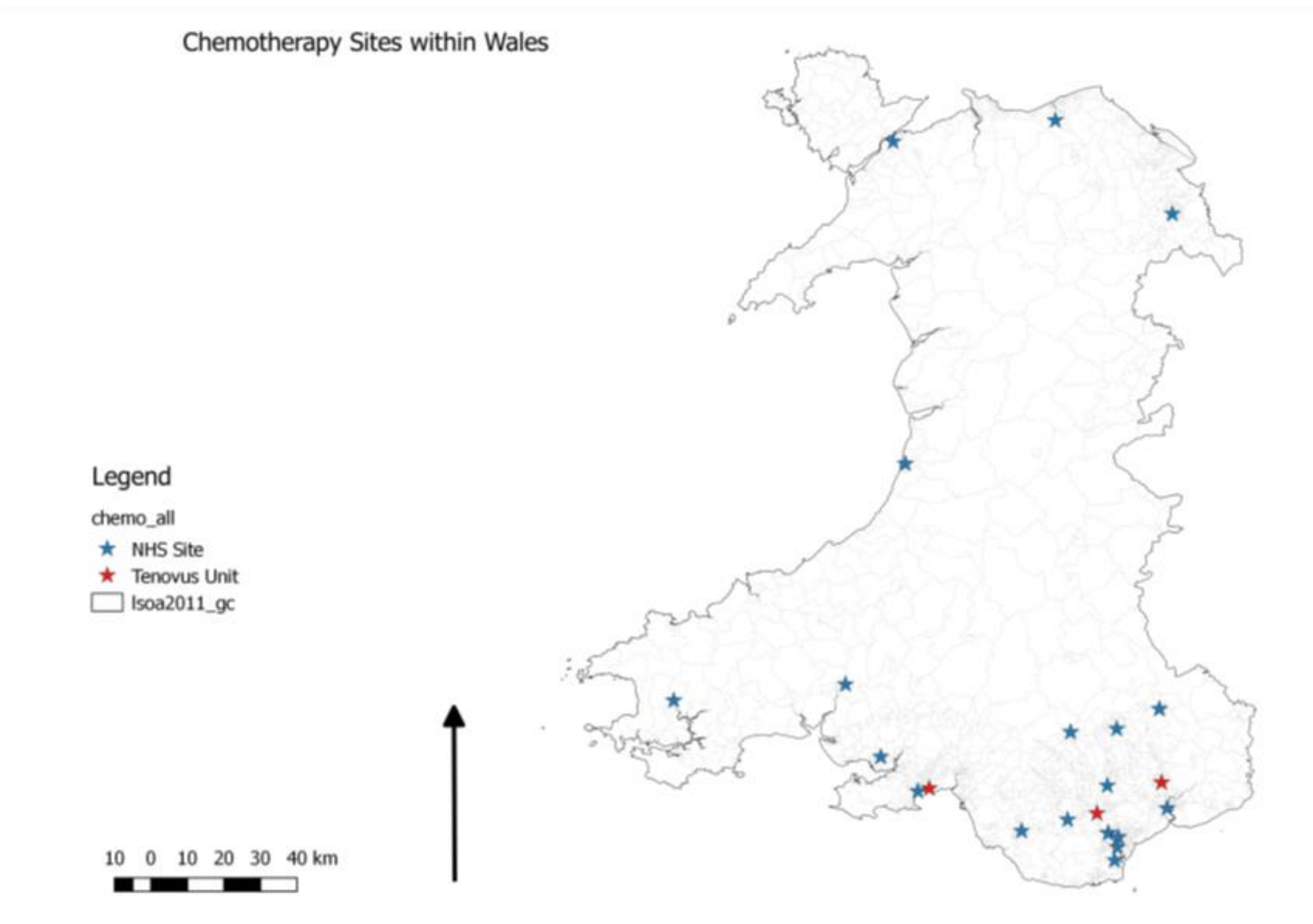


Figure 6-4 Chemotherapy sites within Wales

Figure 6-5 shows the E2SFCA for current NHS chemotherapy provision throughout Wales and the cancer rate within each LSOA. This does not include those individuals who travel to England to access care. This is due to the fact that Tenovus mainly operates in Wales; there are certain cancers which require specialist care and people are required to travel considerable distances to receive this. There are people in large parts of Wales who are unable to access any services within 35km, and many more who have very poor levels of accessibility

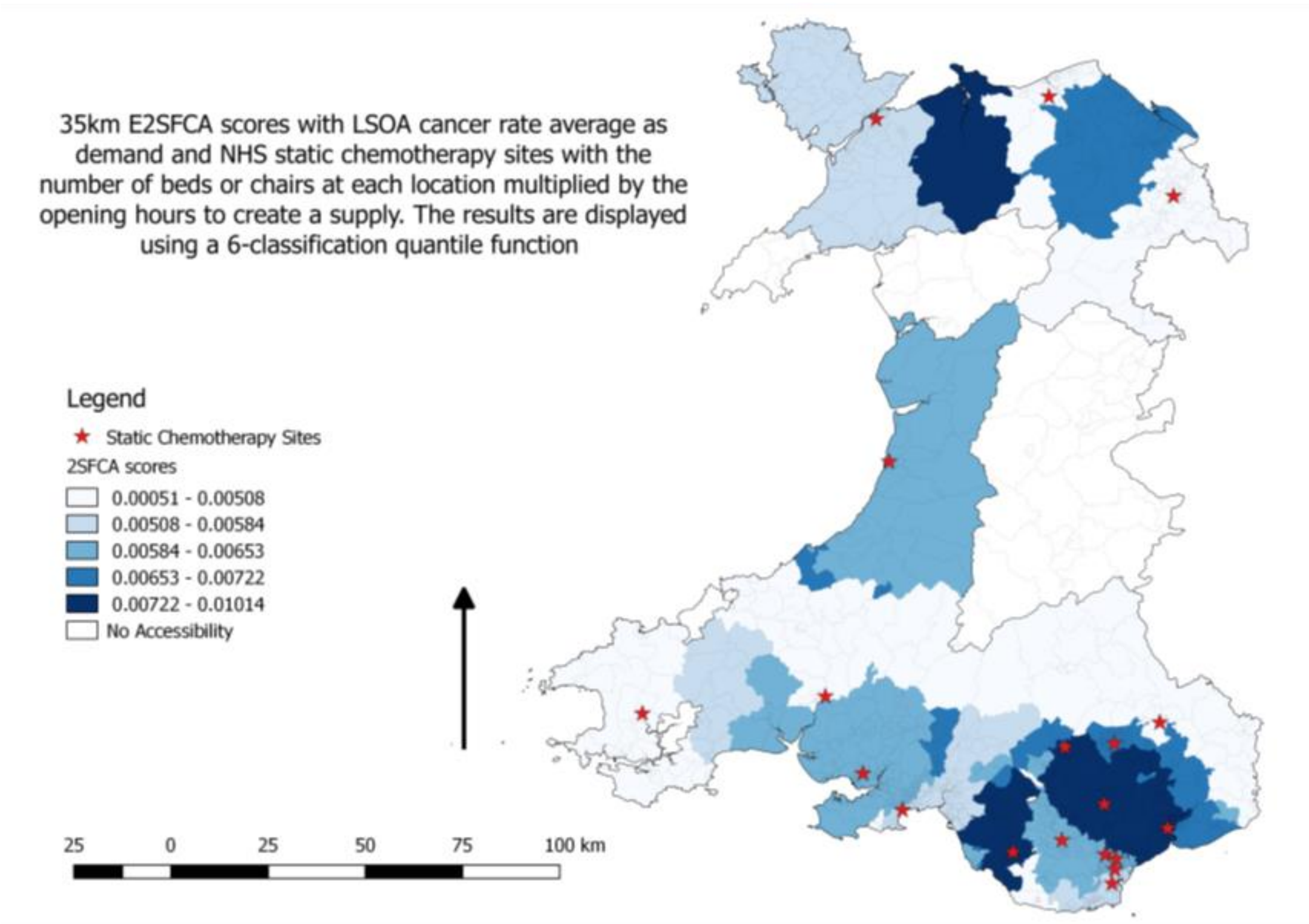


Figure 6-5 35km Linear Distance Decay 2SFCA of the static chemotherapy sites using LSOA cancer rates as demand and a supply of the available hours of each bed/chair per week

Figure 6-6 shows the same accessibility calculations with the Tenovus mobile units included. The current positioning of the mobile units seems to have had very little effect on accessibility scores. This is due to the amount of extra provision they provide. Within the large hospitals the chemotherapy units are able to service a large volume of people. Singleton Hospital provides a supply of 1,800 per week (available hours for each bed/chair) and Royal Gwent Hospital 1,040 whereas a mobile unit can only provide 56. To maximise the potential accessibility gains it was therefore important to address some of the gaps in provision with the types of mobile services offered by Tenovus.

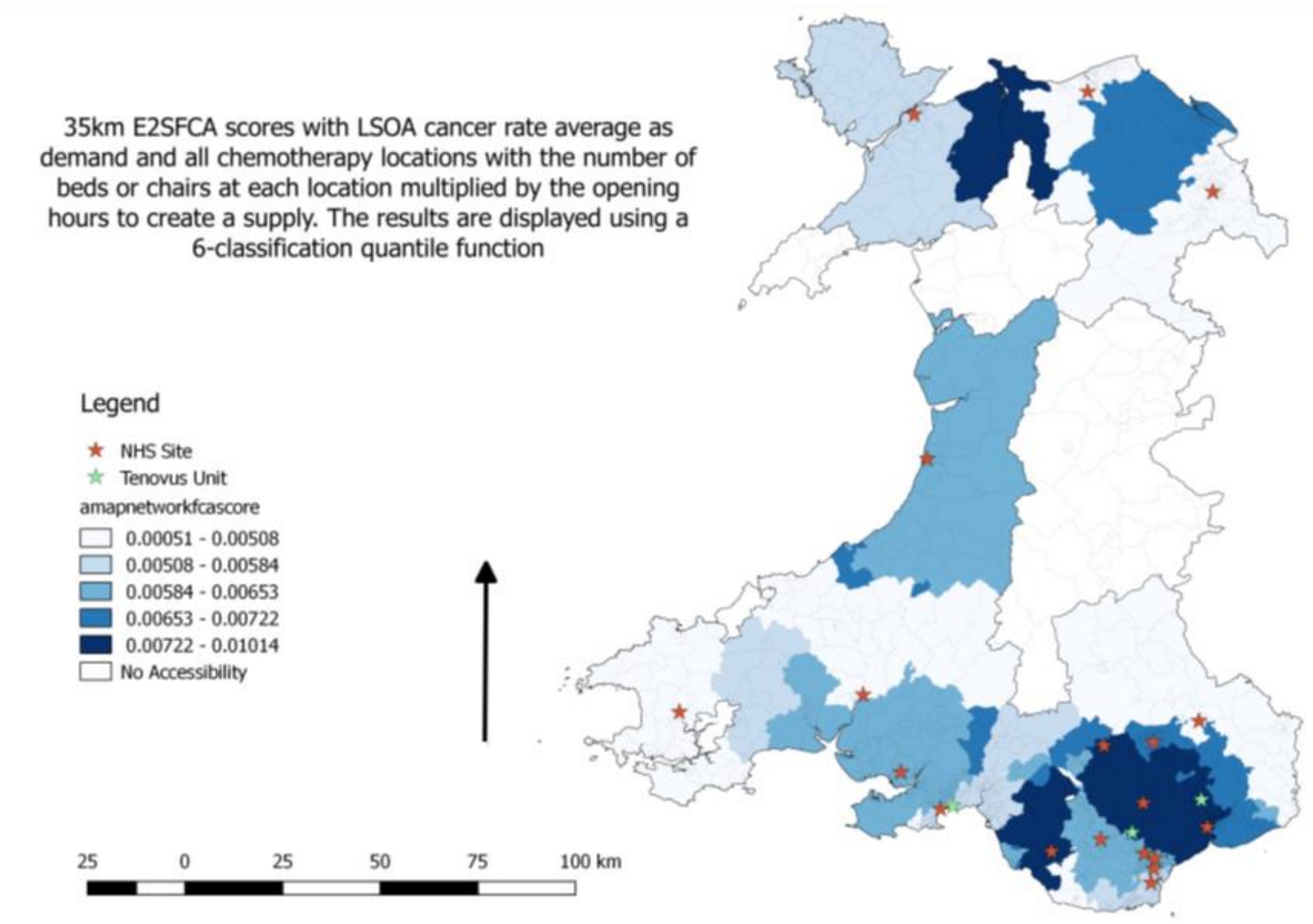


Figure 6-6 35km Linear Distance Decay 2SFCA of the static chemotherapy sites (orange stars) and the Tenovus mobile unit locations (green stars) using LSOA cancer rates as demand and a supply of the available hours of each bed/chair per week

By incorporating a unit at the site of the supermarket in Newtown (Figure 6-7) it is possible to increase the accessibility for individuals living in central Wales dramatically, and using LSOA cancer rate as a demand variable it suggests that the unit could go to the location more than once a week (Figure 6-8).

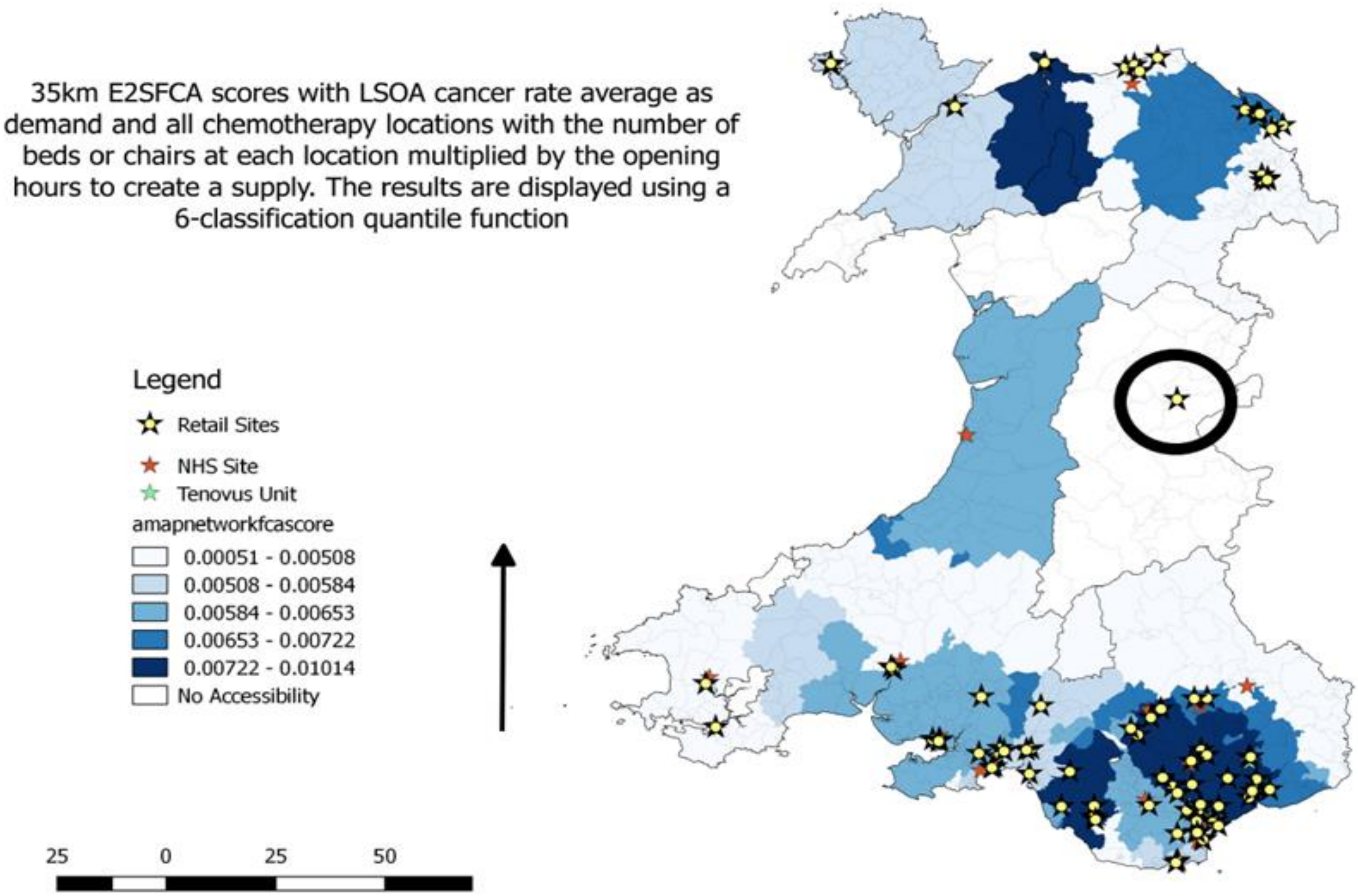


Figure 6-7 E2SFCA and the retail site locations highlighting Newtown, Powys

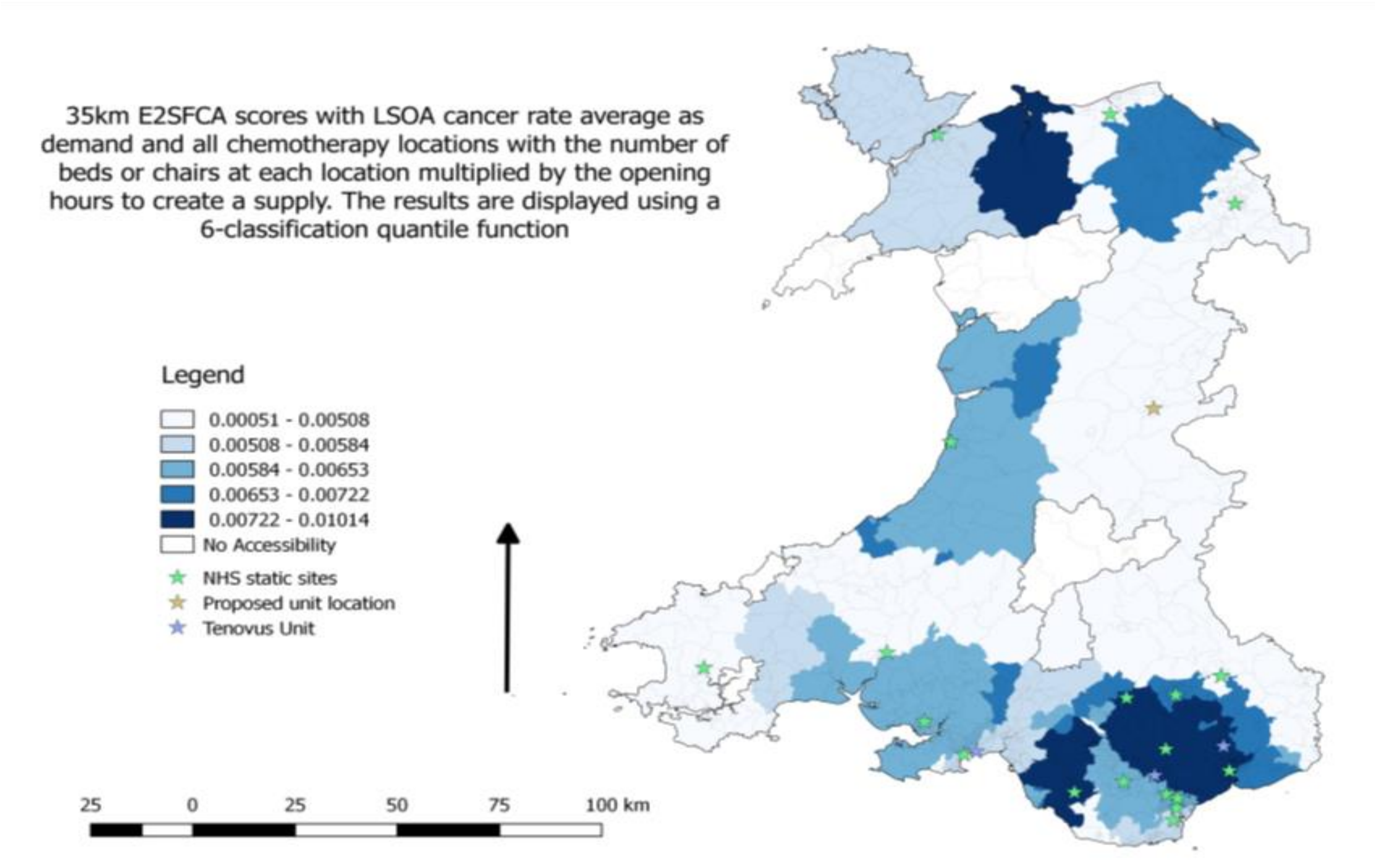


Figure 6-835km Linear Distance Decay 2SFCA of the static chemotherapy sites, Tenovus mobile unit locations and the proposed supermarket in Newtown, Powys using LSOA cancer rates as demand and a supply of the available hours of each bed/chair per week

Figure 6-9 shows the retail units overlaid on the current chemotherapy accessibility map, there is an area of low accessibility surrounding Holyhead which could be improved with additional resource.

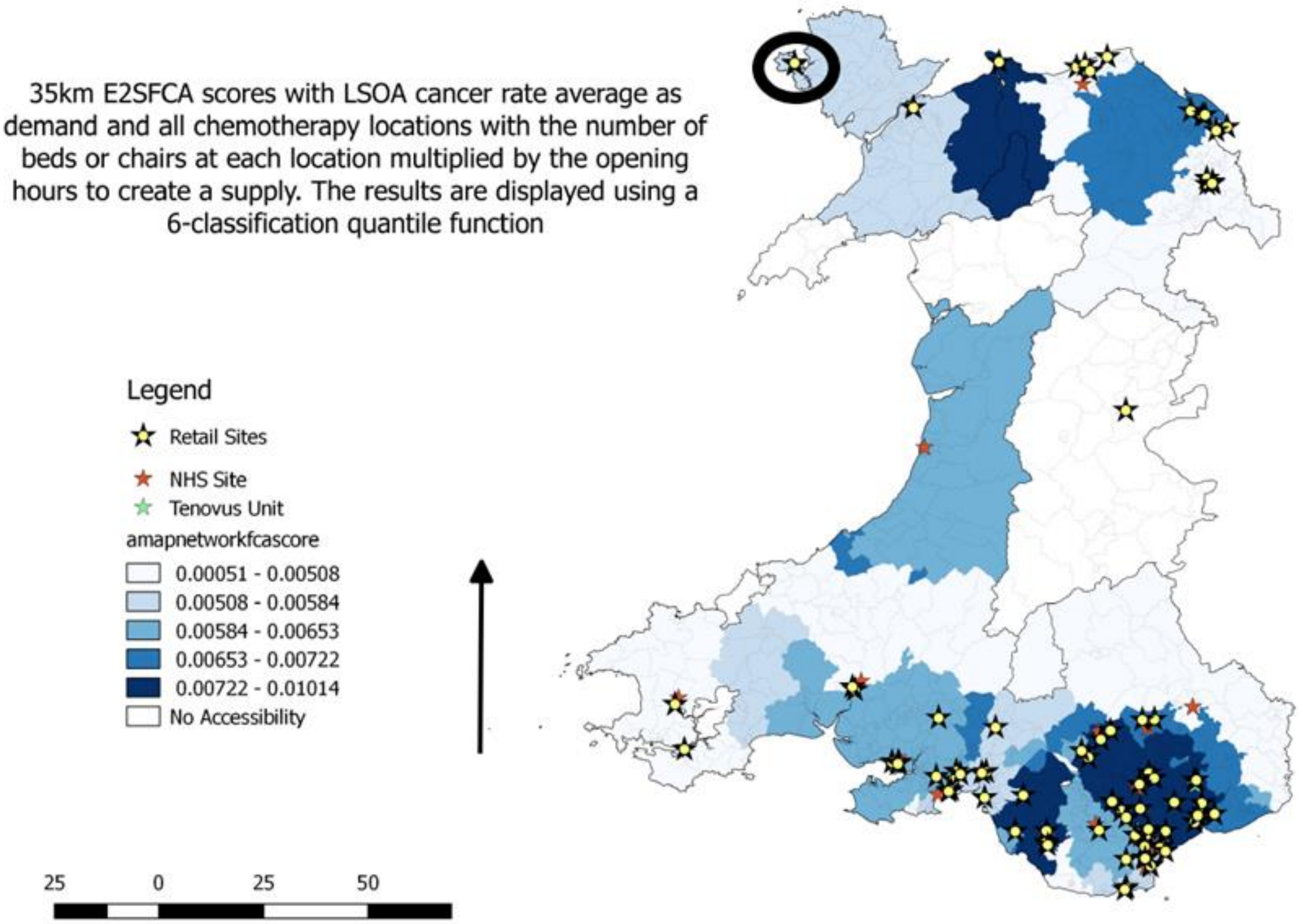


Figure 6-9 E2SFCA and the retail site locations highlighting Holyhead

Figure 6-10 shows E2SFCA accessibility scores when the Holyhead location is serviced by a Tenovus unit, and it makes quite a large difference to the level of accessibility in the north west of Wales. This is another promising potential location for a unit to be positioned as it maximises the effect that the unit has on the surrounding area. By allowing this location to service the chemotherapy requirements of the area it frees up NHS capacity at other locations.

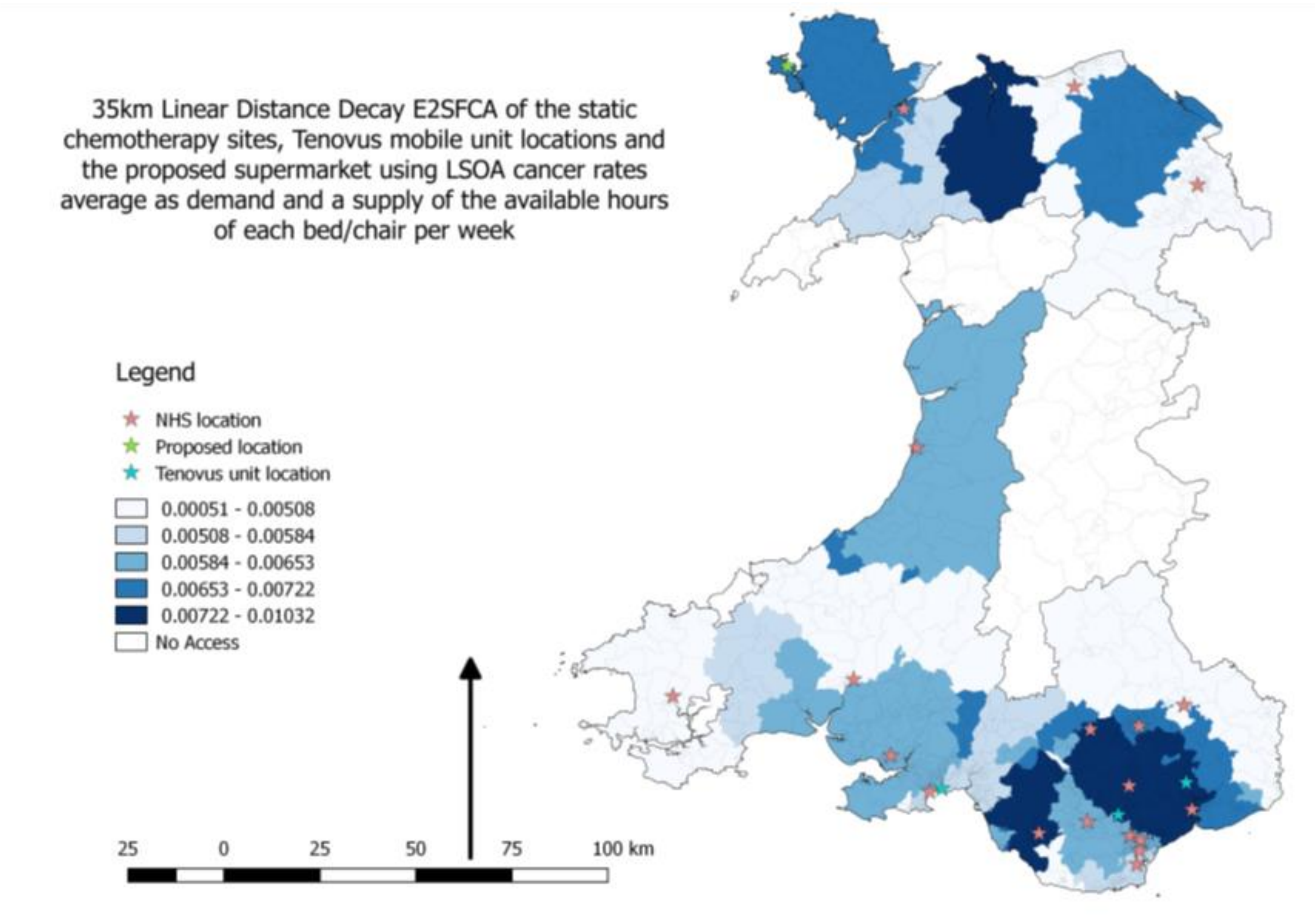


Figure 6-1035km Linear Distance Decay 2SFCA of the static chemotherapy sites (orange stars), Tenovus mobile unit locations (blue stars) and the proposed supermarket in Holyhead (green star)

There is an area of patchy accessibility in the Pyle area of South Wales (Figure 6-11). This could potentially be rectified by adding additional resource from Tenovus. One issue with this part of Wales is that it has a comparatively high population and good transport links with the capital city of Cardiff.

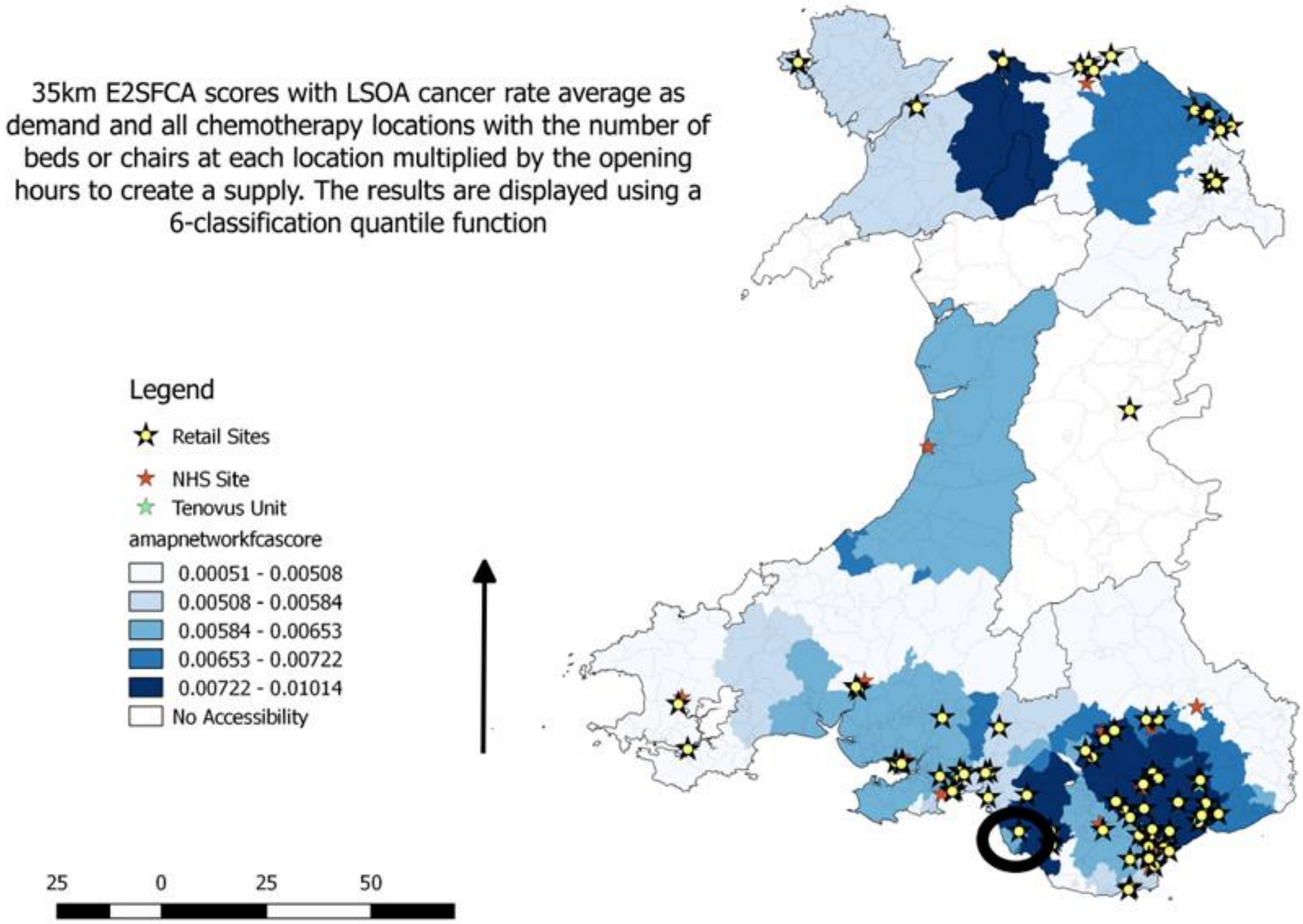


Figure 6-11 E2SFCA and the retail site locations highlighting Pyle

Adding a mobile unit in the Pyle (Figure 6-12) area just West of Cardiff and highlighted in figure 6-11 has very little effect on the surrounding area. There is a boost to the LSOA affected, but it has a limited effect on the surrounding area. This is another important situation where it is possible to show that there is very little benefit in adding additional resources to the area. It is useful for Tenovus to be able to understand not only the positive effects of their presence but also the times in which additional resource has a limited effect in the surrounding area.

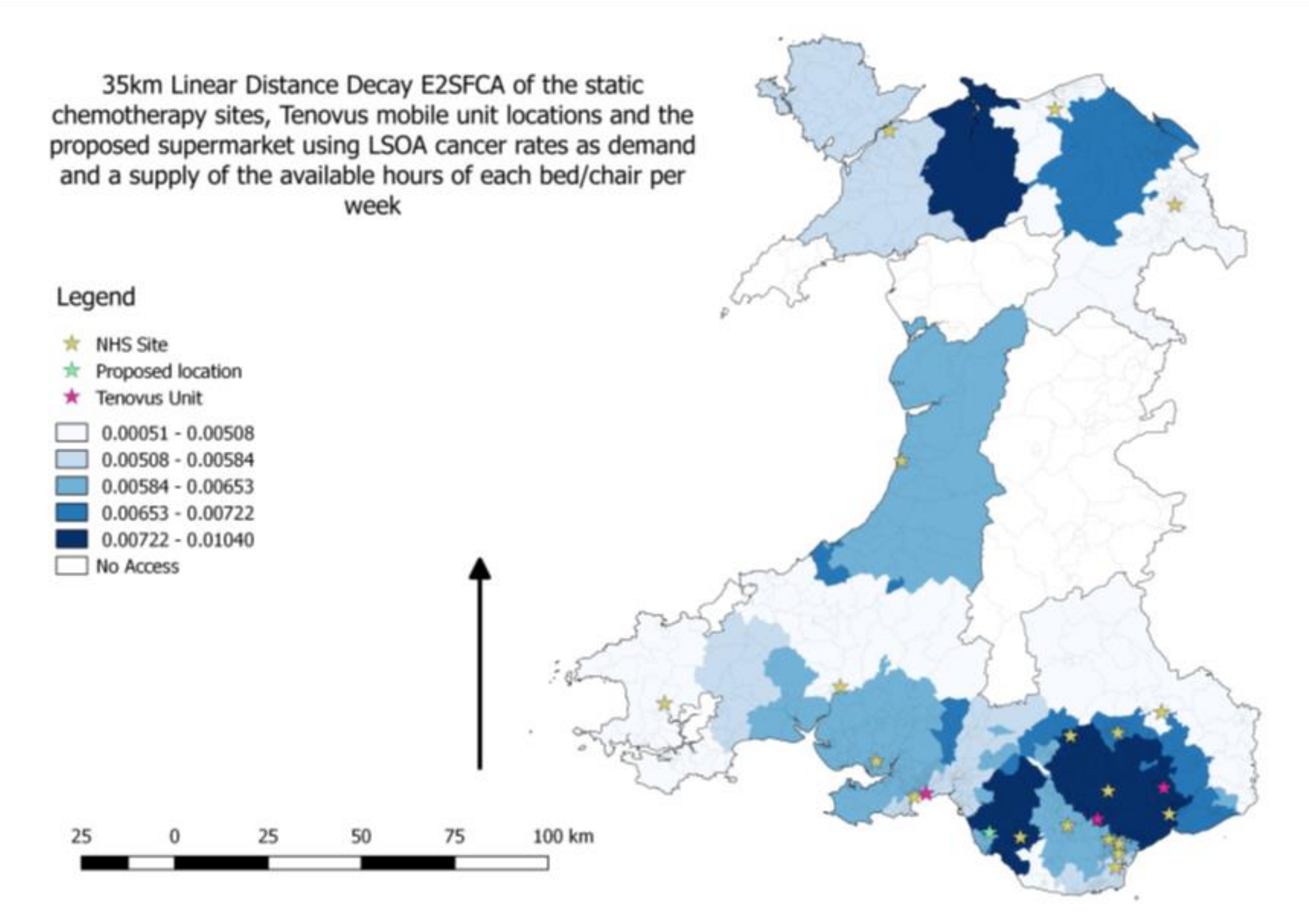


Figure 6-12 35km Linear Distance Decay 2SFCA of the static chemotherapy sites (red stars), Tenovus mobile unit locations (yellow stars) and the proposed supermarket in Pyle (Blue star)

Using chemotherapy supply data that has been gathered, it is possible to get an overview of current chemotherapy provision in Wales. The purpose of this exercise is to show the effects of adding additional resources to a new location.

The case study showed that Tenovus was having a minimal effect in terms of the overall patterns of accessibility to Chemotherapy services in Wales based on the current locations of the mobile units. It also demonstrated that there were some potential locations which could improve access to these services for potential patients in other areas of Wales. It must be noted that it is not just Tenovus that decide where their resources should be deployed. Currently, the local health boards decide whether they would like Tenovus in their area; and they are also governed by how far the NHS staff that provide the treatments are able to travel.

The location with the largest potential effect is in Powys, many of whose patients are travelling across the border to England to access cancer services. They have no chemotherapy provisions at all in the health board; as such it would be difficult for Tenovus to provide this service as there would be no staff to operate the unit. With this in mind it is still important to acknowledge the effects a potential supply in these locations could have. This would provide Tenovus with the ability to speak to Powys health board with an informed opinion of the difference they could make if patients were to be offered provision within Wales.

6.3 Visualisation of results

Several visualisations have been created for Tenovus Cancer Care to give them further understanding of current chemotherapy provisions within Wales. These visualisations have been created utilising the 2SFCA tool and QGIS. Having the ability to create visualisations enables Tenovus to gain an understanding of current and modelled levels of provision and present the data to interested stakeholders in a variety of forms.

There were also requests to create maps of other services such as the choirs they run for those affected by cancer, their outreach services and their charity shops. It has been important to visualise the data so that the people at Tenovus could understand the potential of the 2SFCA tool and help create a dialogue between key stakeholders in the project.

Figure 6-13 shows the different chemotherapy sites in Wales. An E2SFCA method was used with cancer rate of each LSOA as the demand, and the number of beds or chairs at each location multiplied by the opening hours, to create a supply with a 35km catchment area. The 35km catchment area was chosen as Tenovus user data showed the vast majority of users travel within this distance (See section 4.3.3). The results showed that there is a large part of Wales with no access to chemotherapy and access outside of the larger metropolitan areas of Cardiff, Aberystwyth, Swansea, Bangor and Llandudno is low. When looking at chemotherapy provision across Wales it was possible to identify areas which could be improved to influence overall patterns of accessibility.

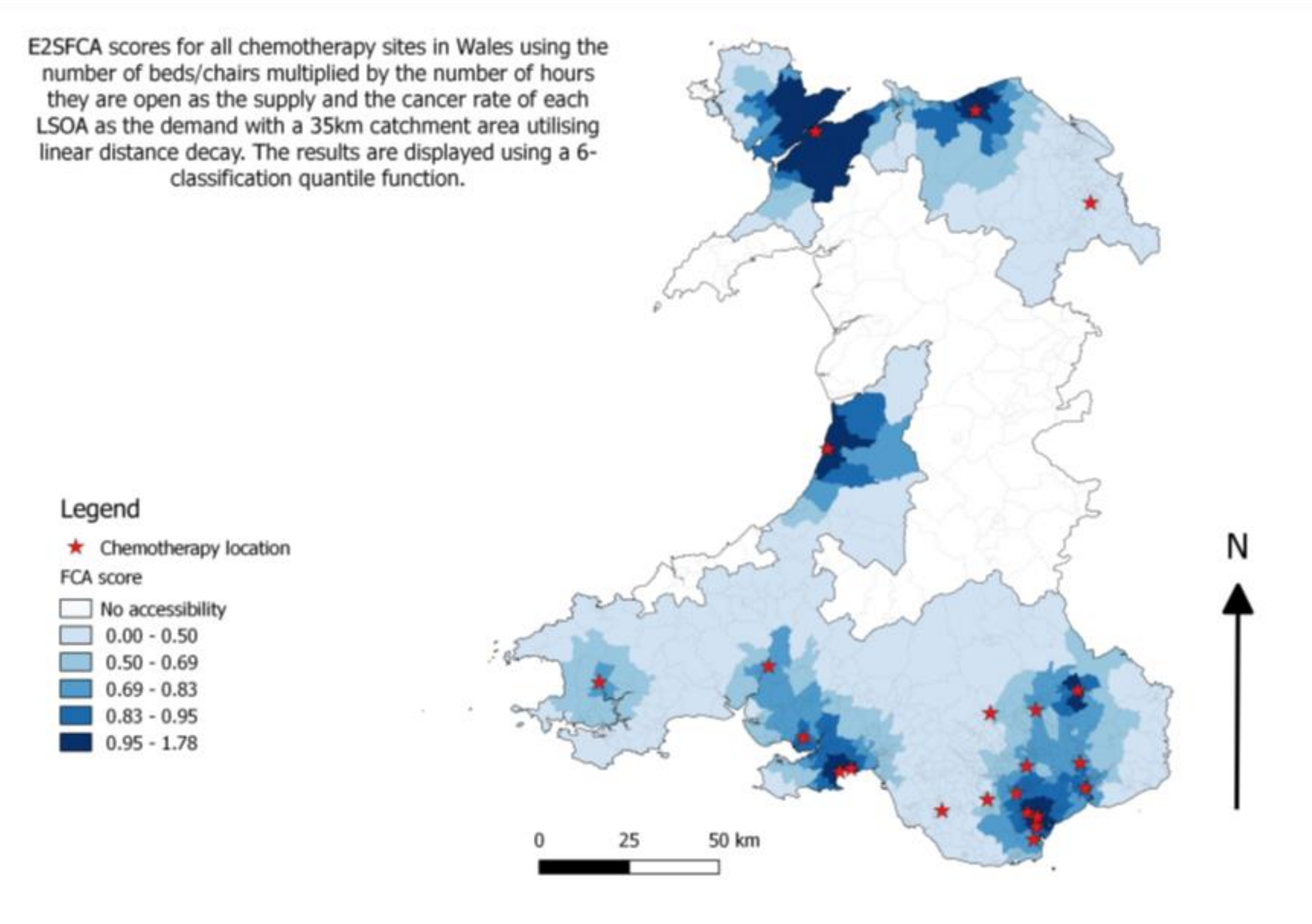


Figure 6-13 Chemotherapy provisions in Wales E2SFCA

Figure 6-14 shows different chemotherapy sites in Wales. An E2SFCA method was used with the total population of each LSOA as the demand, and the number of beds or chairs at each location multiplied by the opening hours, to create a supply with a 35km catchment area. This map highlights that apart from a spike in the Bangor region there is reasonably low accessibility to cancer services across Wales.

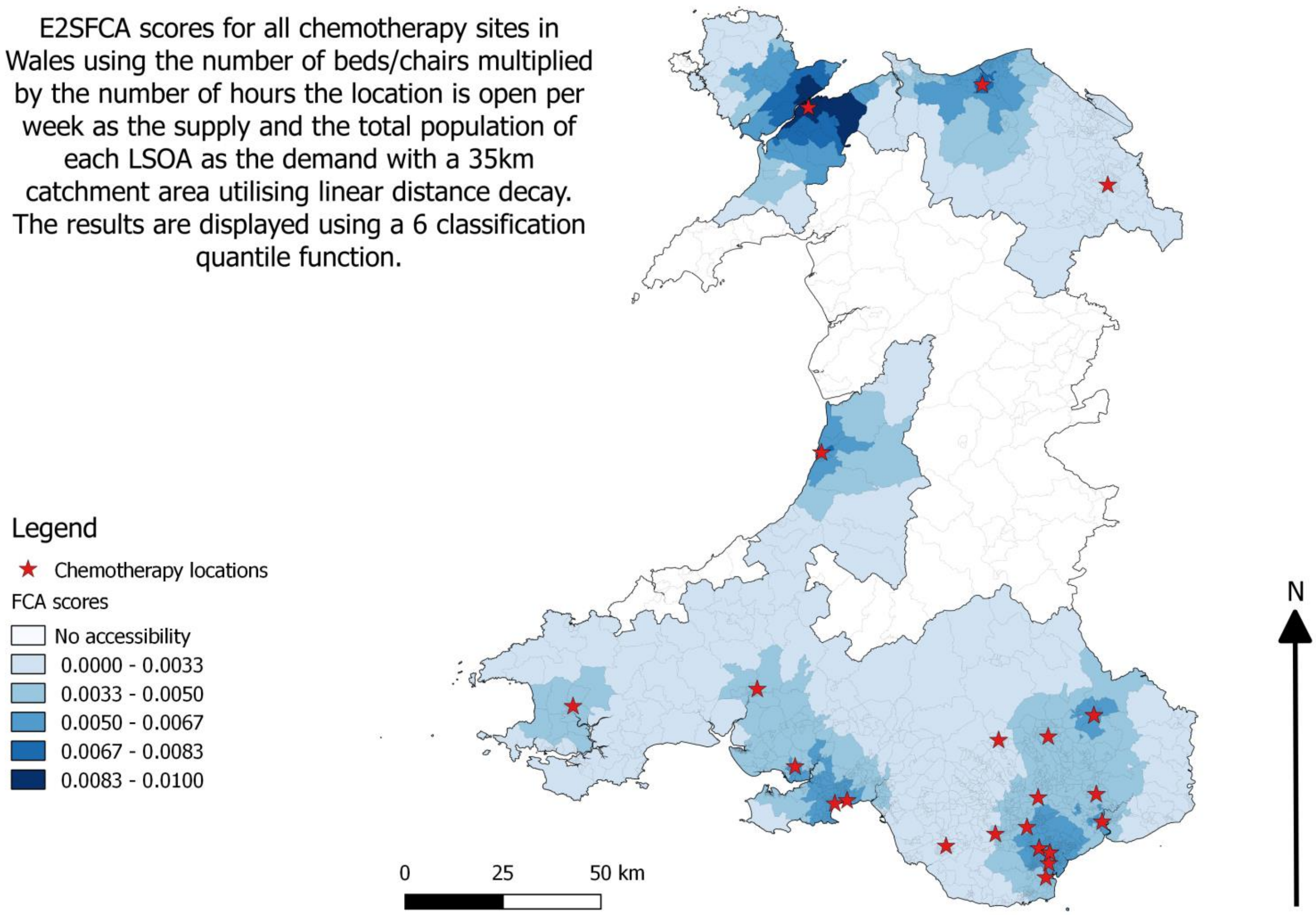


Figure 6-14 E2SFCA Chemotherapy sites supply total population demand

Figure 6-15 shows Tenovus choir sites across Wales. 2SFCA was completed using LSOA cancer incidence as the demand with a 40km catchment area. The visualisation highlights that although the majority of the choir sites are in the South East it is not saturated, and although there are relatively few sites in mid and west Wales general access to these choirs is comparatively high.

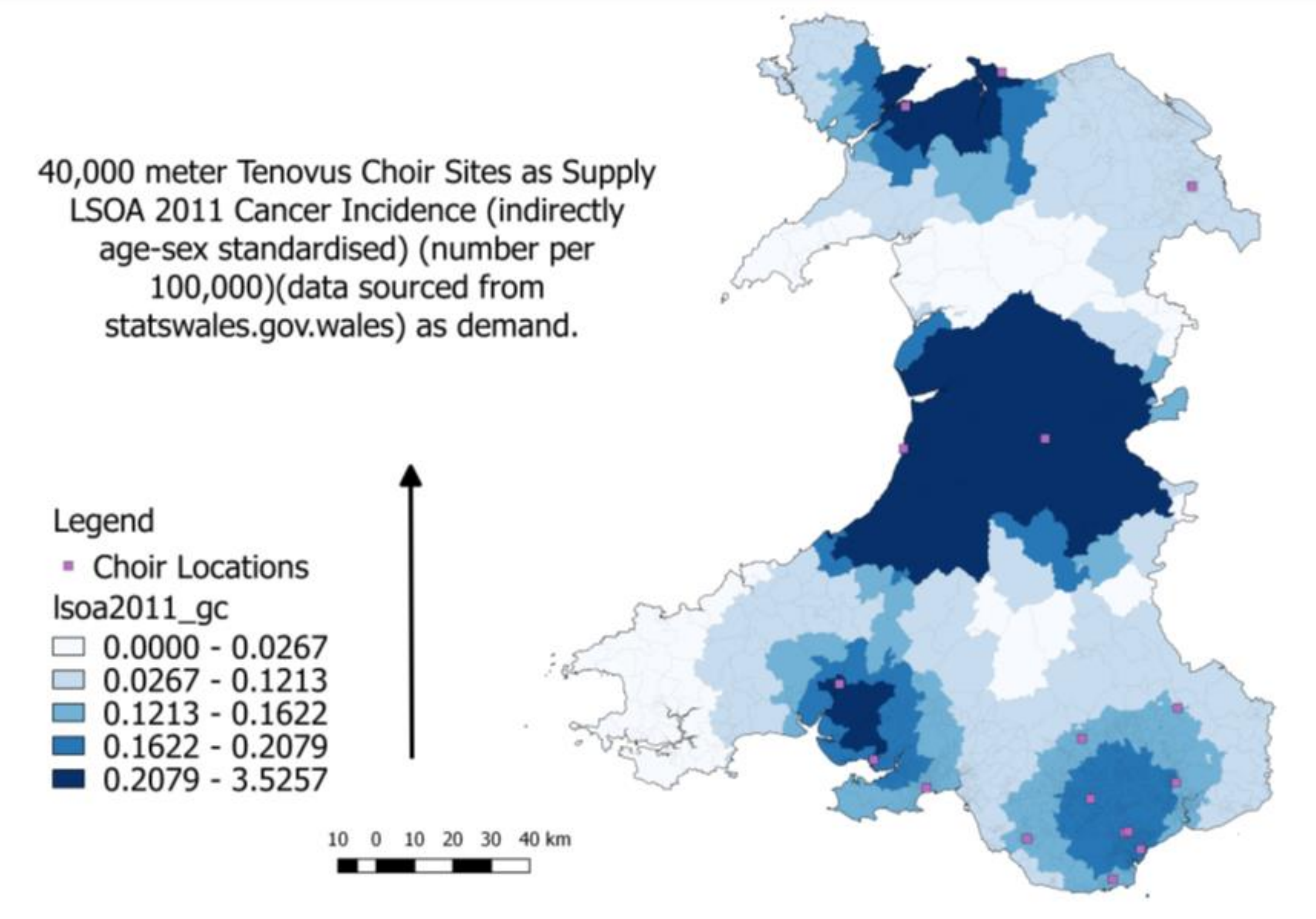


Figure 6-15 Tenovus Choir sites LSOA cancer incidence 2SFCA

Figure 6-16 shows Tenovus outreach sites across Wales. 2SFCA was completed using LSOA cancer incidence as demand with a 40km catchment. The data shows that there is a large proportion of Wales with no access to the service, and although the south east has a high proportion of the sites it is not saturated. By using the cancer incidence rate, it was possible to gain a richer understanding of the users of the service than by using the total population as whole.

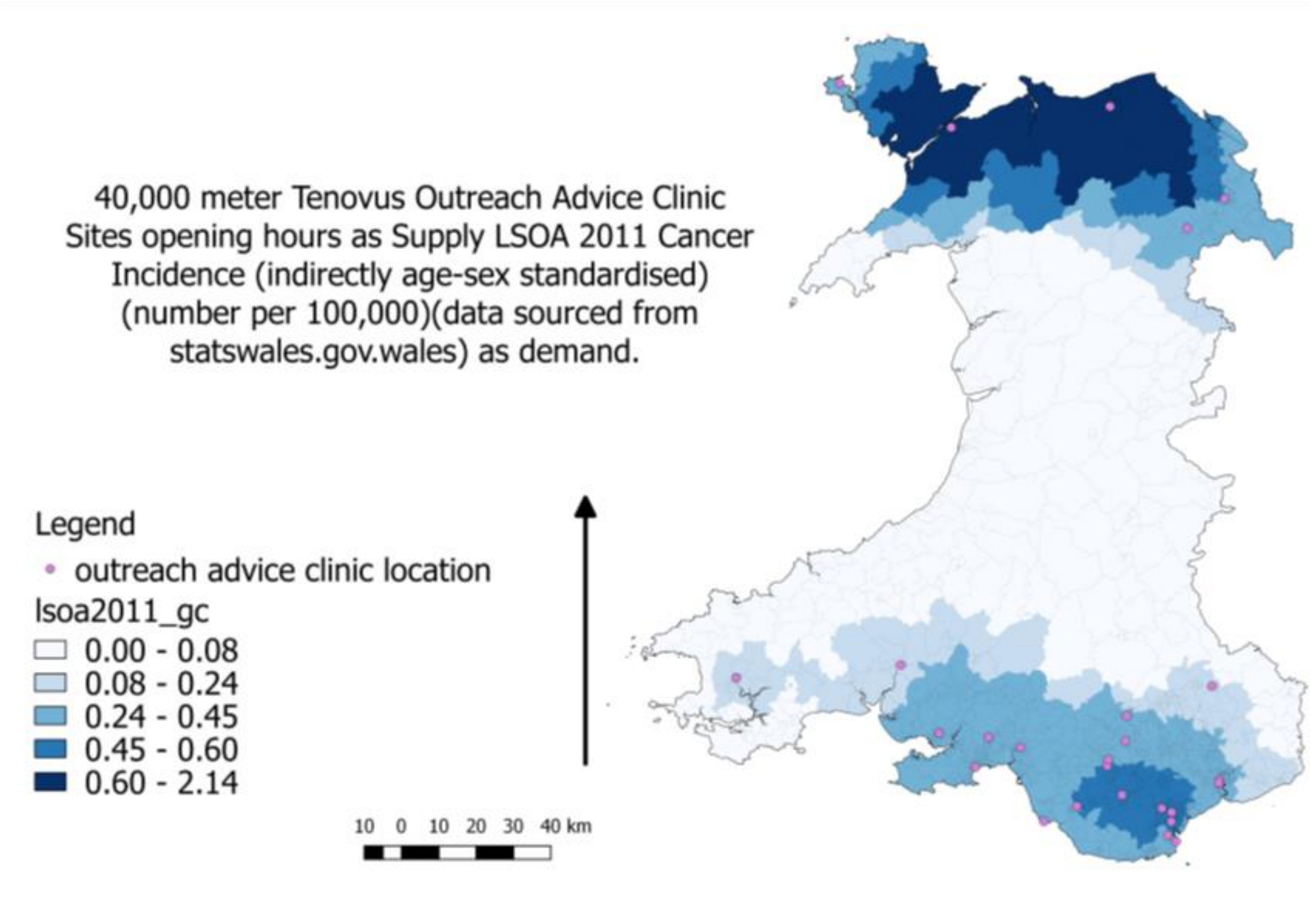


Figure 6-16 2SFCA Tenovus Outreach Clinics locations LSOA cancer incidence 2SFCA

In Figure 6-17 it is possible to see locations of Tenovus charity shops. 2SFCA was used to show accessibility, although this is not the best measure for retail units. For instance, much of the footfall in west Wales may come from holiday makers, and people travel to shop differently from how they access healthcare provisions. Total population was used as the demand and a 40km catchment was employed. The map shows that although a large proportion of the sites are in the south east, the area is not saturated and there is very good access in the west of Wales.

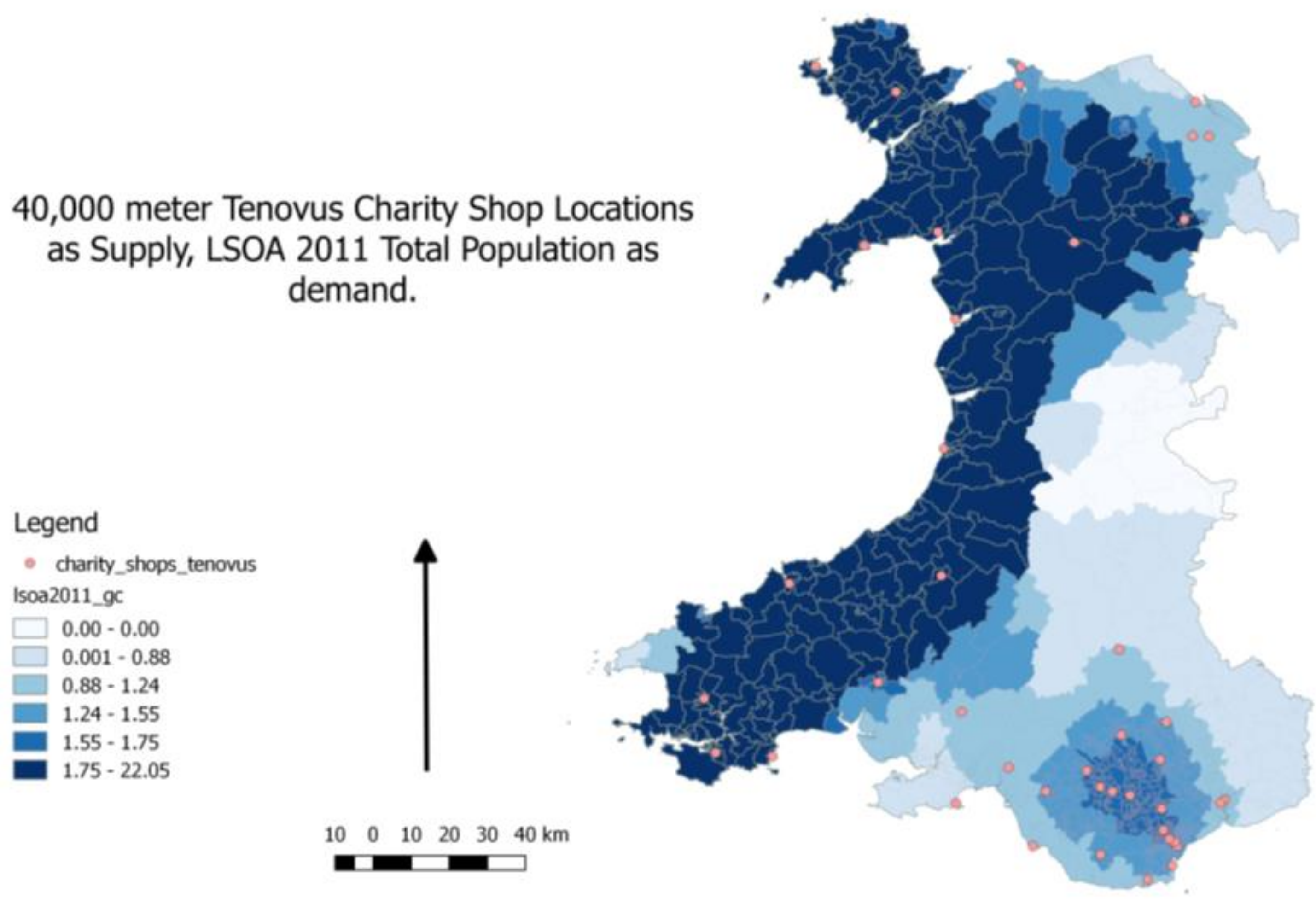


Figure 6-17 Tenovus Charity Shop locations LSOA total population 2SFCA

Figure 6-18 shows all Tenovus sites across Wales and volunteer locations surrounding them. This visualisation was used to see trends in whether volunteers spiked around the service provision. It is clear that the coast of Wales is fairly well serviced but mid Wales is still reasonably under-represented.

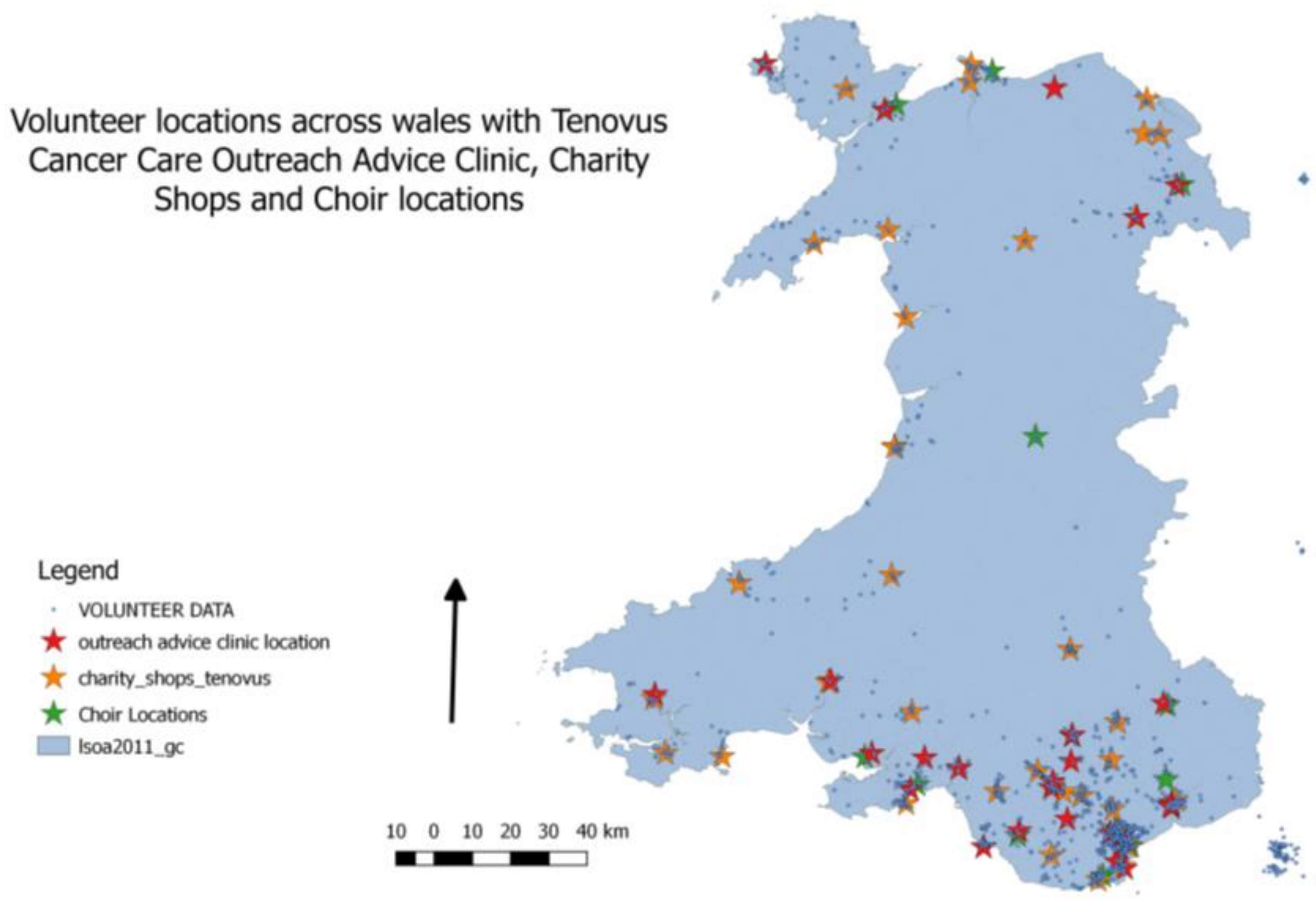


Figure 6-18 Tenovus sites across Wales and volunteer data

These visualisations have enabled Tenovus to have a richer understanding of their services, and enabled key stakeholders of the project to understand the potential of the tool. Creating these and explaining them to Tenovus employees initiated a dialogue and enabled more feedback to be presented. These maps could be created by Tenovus to show their supporters the work they are doing, or they could be used to help show partners' needs in accessibility throughout Wales in order to gain more funding

6.4 Data

6.4.1 Scale comparison

The following images look at the differences in results when E2SFCA and linear E2SFCA have been calculated at different scales. The calculations were performed at 10km and at 40km across OA, LSOA and MSOA. The NHS defines the LSOA as follows, “Lower Layer Super Output Area (LSOA) is a GEOGRAPHIC AREA. Lower Layer Super Output Areas are a geographic hierarchy designed to improve the reporting of small area statistics in England and Wales”. MSOA can be described as “A Middle Layer Super Output Area (MSOA) is a GEOGRAPHIC AREA. Middle Layer Super Output Areas are a geographic hierarchy designed to improve the reporting of small area statistics in England and Wales, and are built from groups of contiguous Lower Layer Super Output Areas.” Finally, OAs are the Output Areas that make up the MSOA and LSOAs.

Figures 6-19, 6-20 and 6-21 show the differences between different scales. The maps show the E2SFCA results with a Gaussian distance decay function applied at OA, LSOA and MSOA levels with a 10km catchment area. In general, the maps show a similar pattern with the most detail showing with the OA data followed by the LSOA data as would be expected. This is most prevalent in the built-up areas of south east Wales.

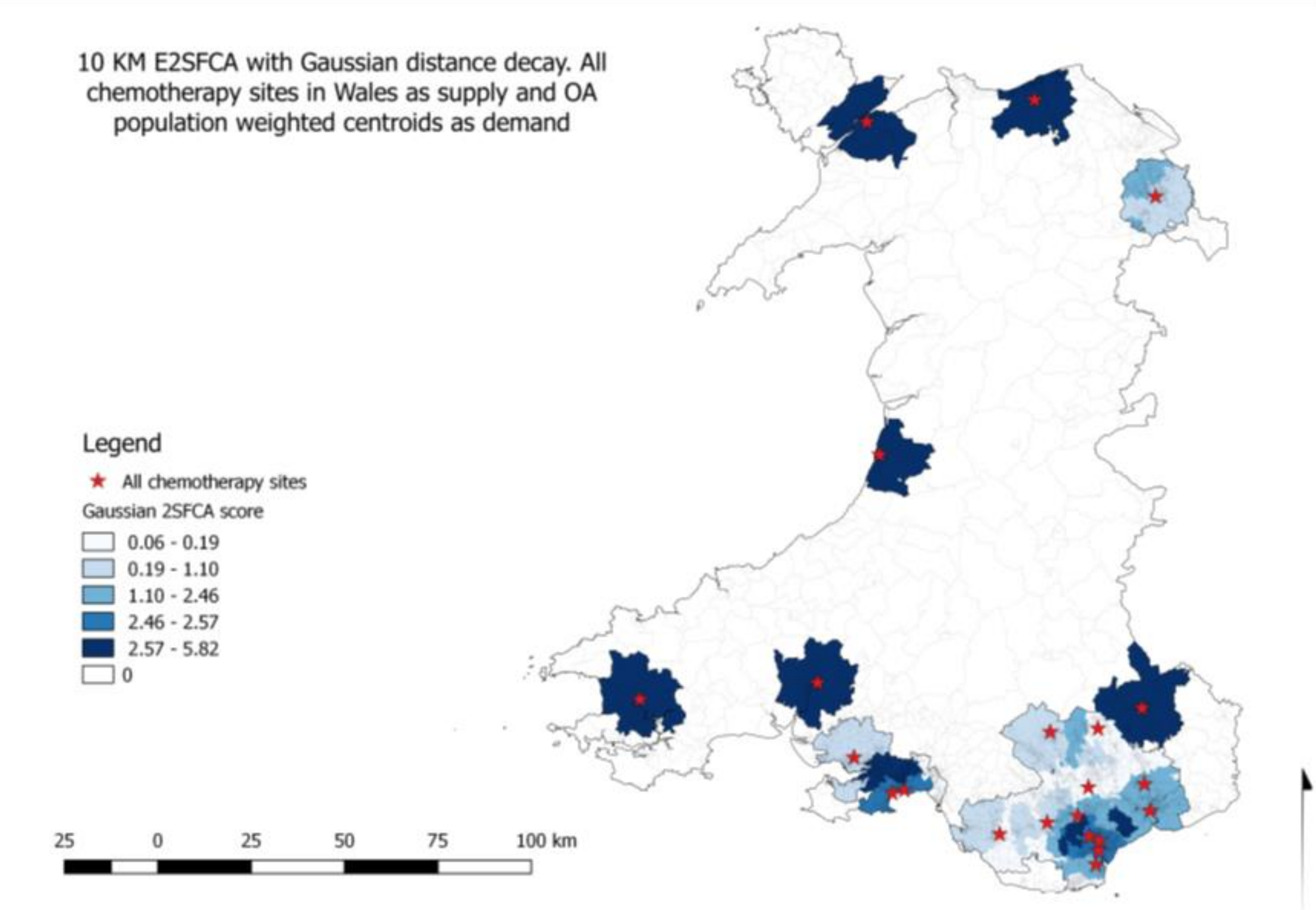


Figure 6-19 10km Gaussian E2SFCA results at OA scale

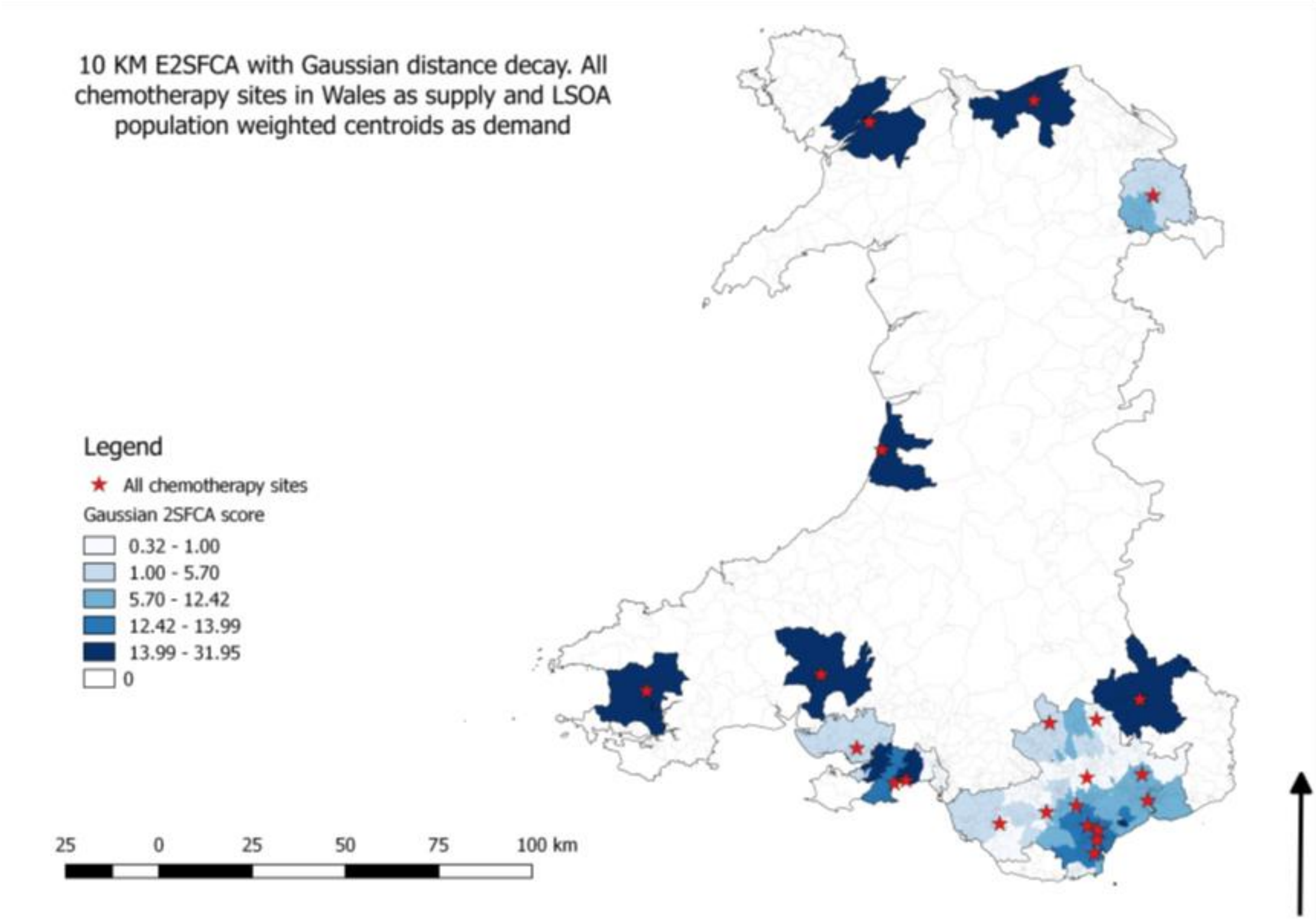


Figure 6-20 10km Gaussian E2SFCA results at LSOA scale

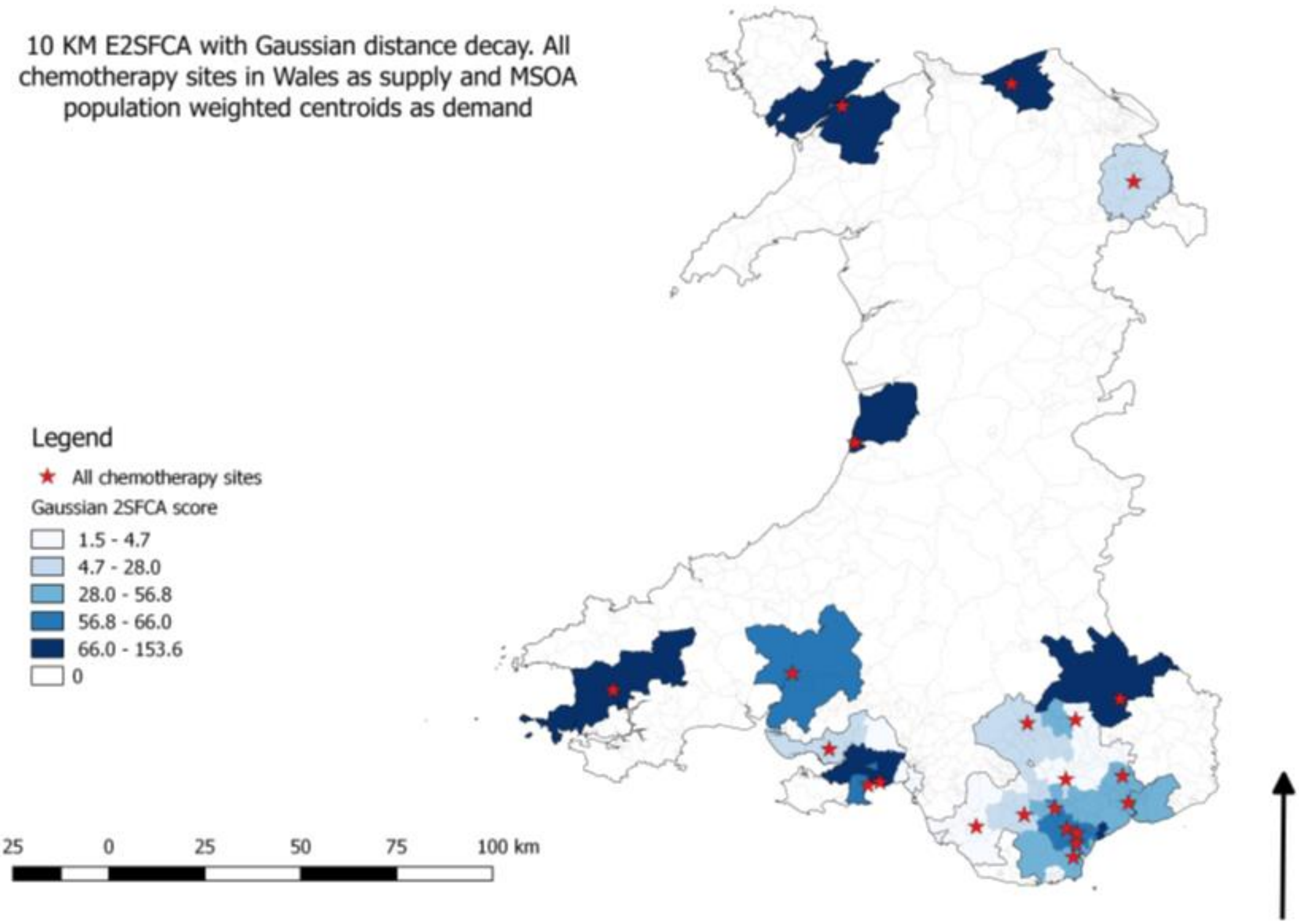


Figure 6-2110km Gaussian E2SFCA results at MSOA scale

Table 6.5 shows that there are many more results for the OA data, as there are many more datapoints, having 5 times that of the LSOA data, and the LSOA being around 5 times more than MSOA. These results helped assist with the visualisation, and were important when investigating data in built up areas where the small area variations could remain hidden with the aggregation of the data.

10km	**Mean**	**Total results**
OA	1.7	6530
LSOA	8.8	1257
MSOA	42.1	266

Table 6.5 10km means and number of results

Figures 6-22, 6-23 and 6-24 investigated the scale effects at a distance more relevant to Tenovus. As shown in section 4.3.3 the furthest distance most Tenovus users travel is within a 35km catchment area and understanding the effect of scale on the data helps in guiding Tenovus to the correct data to use. A similar pattern was apparent as with the 10km results and in the large areas there was more distinction in more built up areas. It also showed slightly different areas due to the size of the output areas.

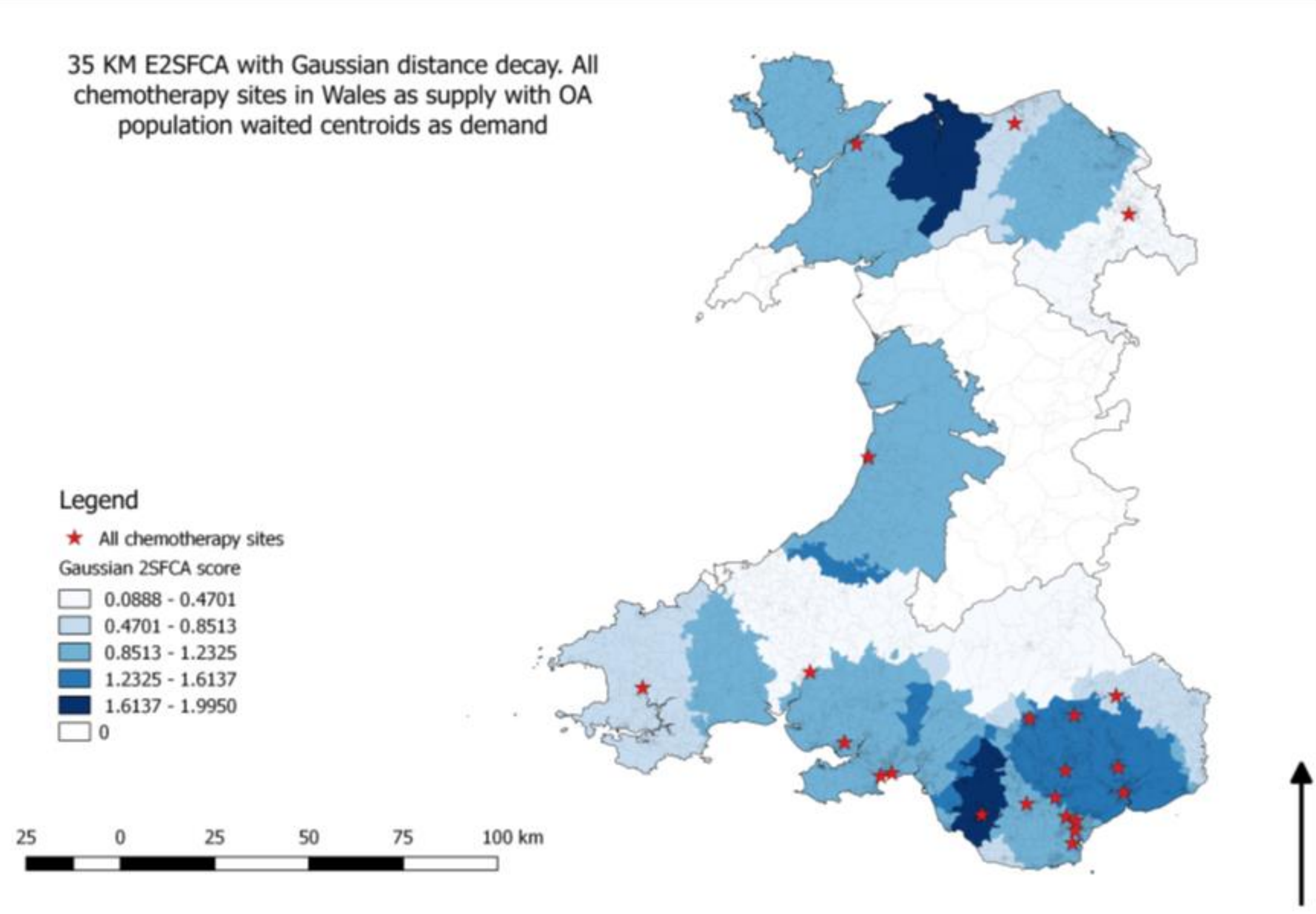


Figure 6-22 35km Gaussian E2SFCA results at OA scale

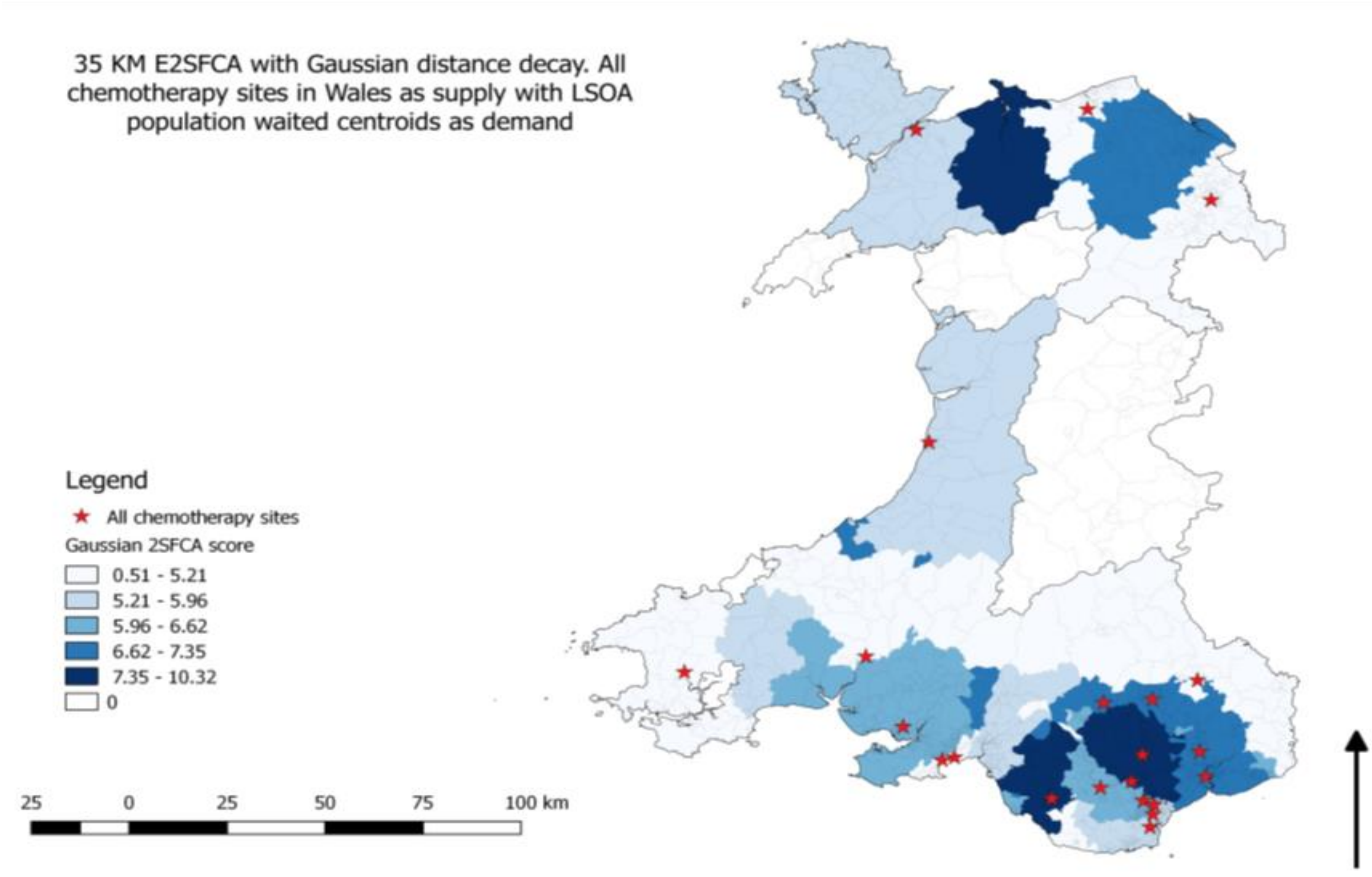


Figure 6-23 35km Gaussian E2SFCA results at LSOA scale

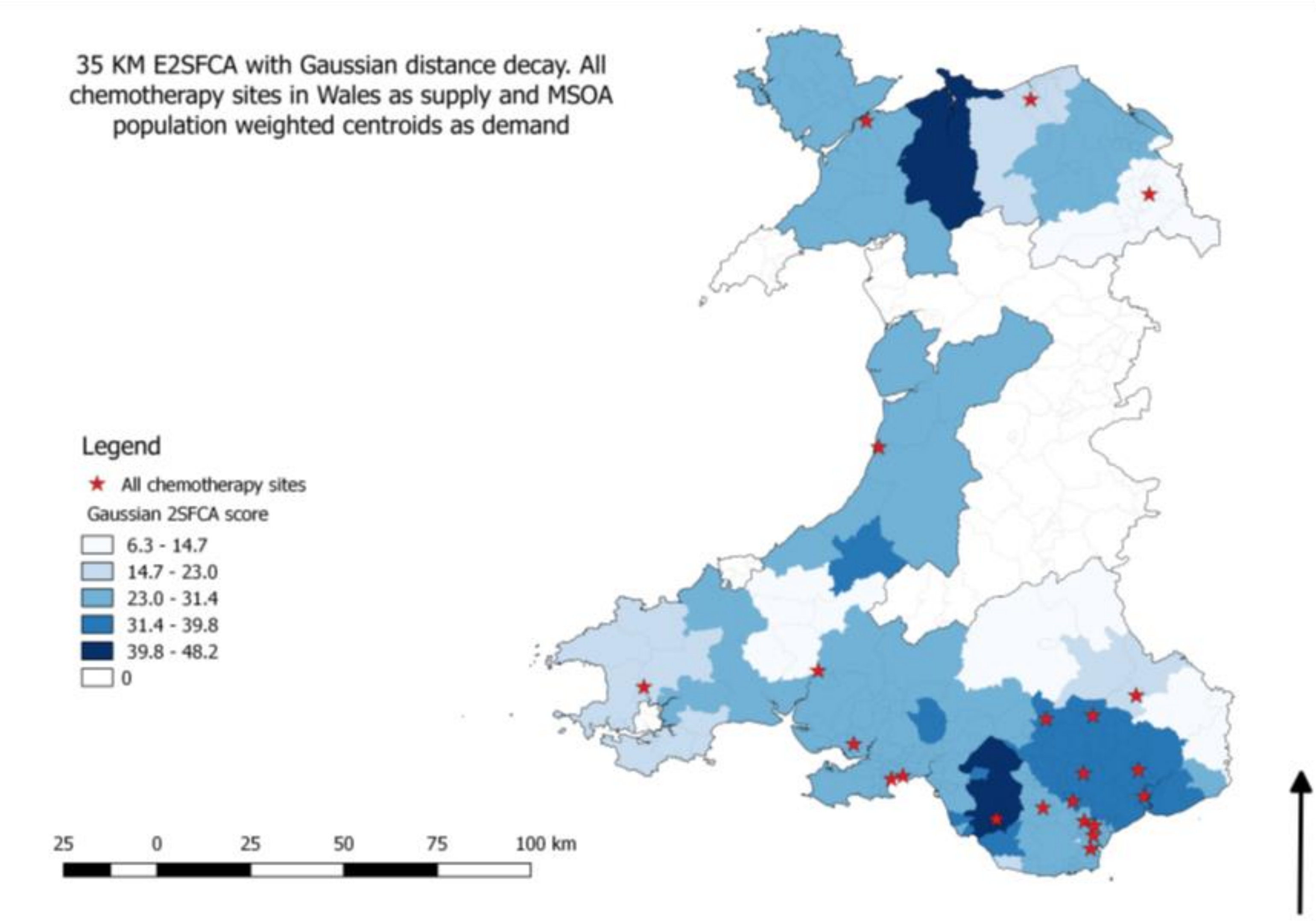


Figure 6-24 35km Gaussian E2SFCA results at MSOA scale

Table 6.6 shows a similar pattern to that at 10km and there are considerable variations between the data sets. This results in a difference in granularity on the maps produced. At this scale the differences seem to be smaller in appearance, and although utilising the OA data would produce a more complete picture, analysis of provision at a variety of spatial scales results in readable maps which can provide valuable insights.

35km	Mean	Total results
OA	1.1	9624
LSOA	6.1	1837
MSOA	28.7	390

Table 6.6 35km means and number of results

Figures 6-25, 6-26 and 6-27 compare scale of a larger catchment size again. At 50km the differences between the data scales had less of an impact as the size of the catchment increased. It still highlighted built up areas in more detail, and this should be considered when designing a study which is predominantly in metropolitan areas.

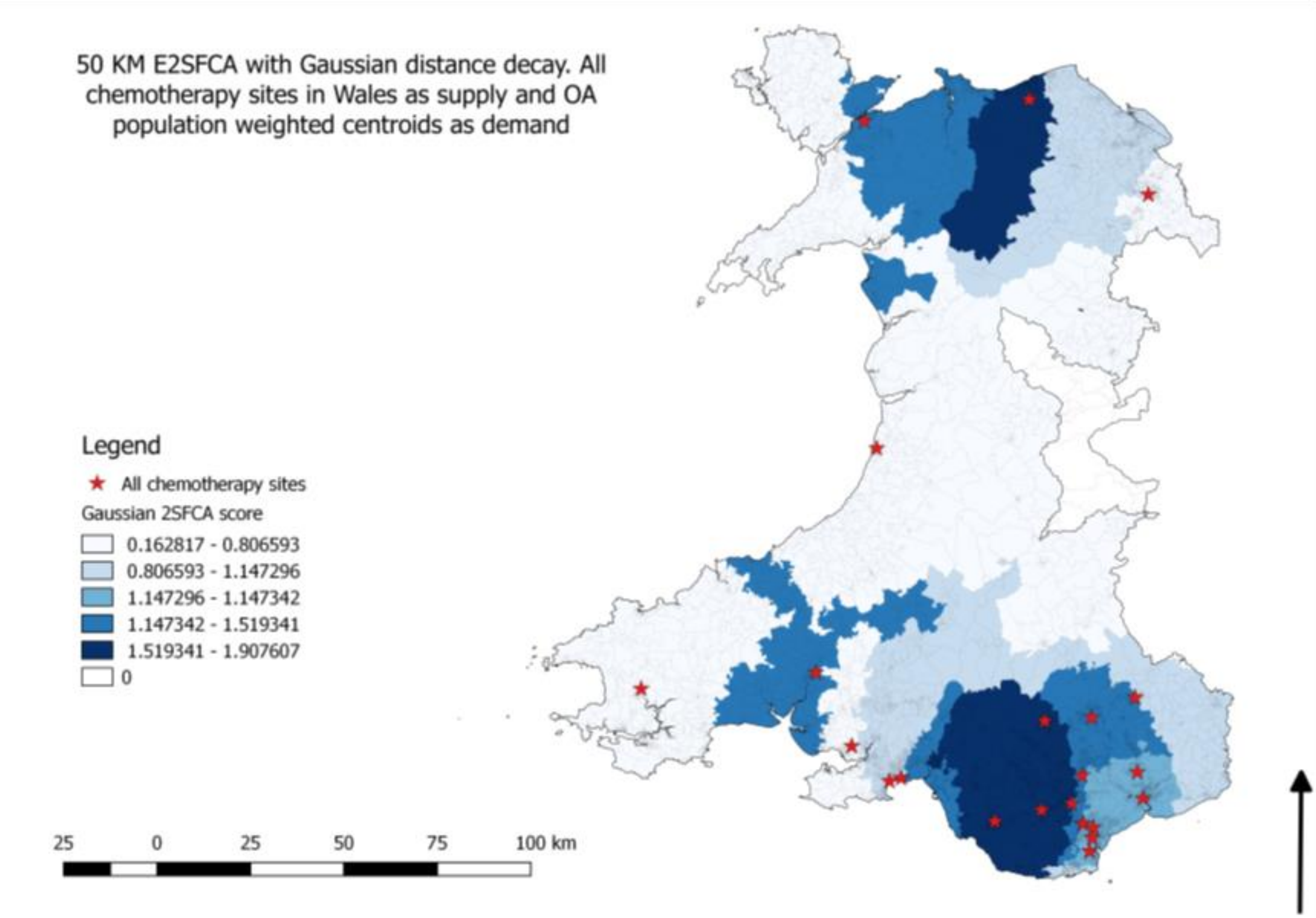


Figure 6-25 50km Gaussian E2SFCA results at OA scale

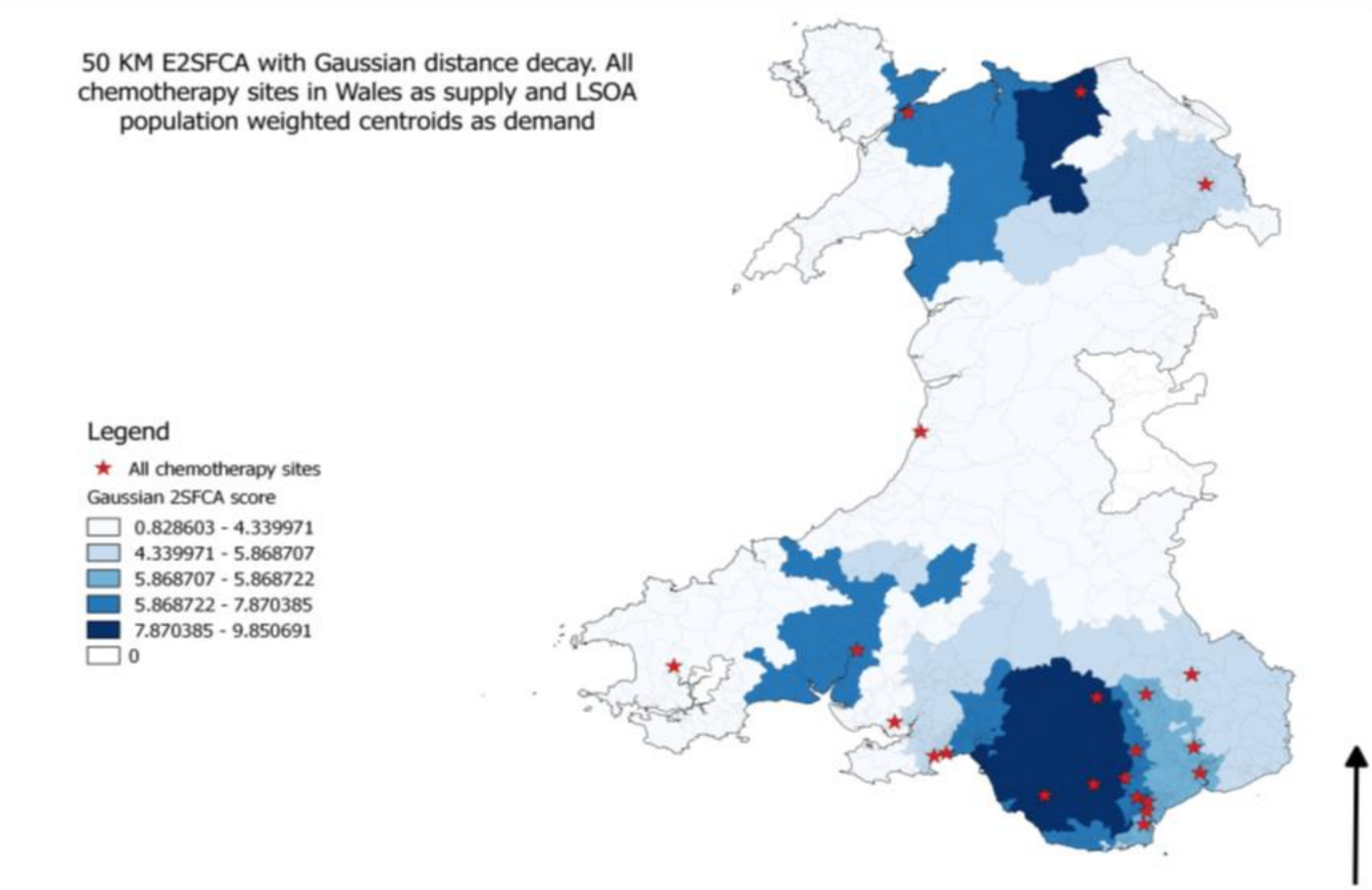


Figure 6-26 50km Gaussian E2SFCA results at LSOA scale

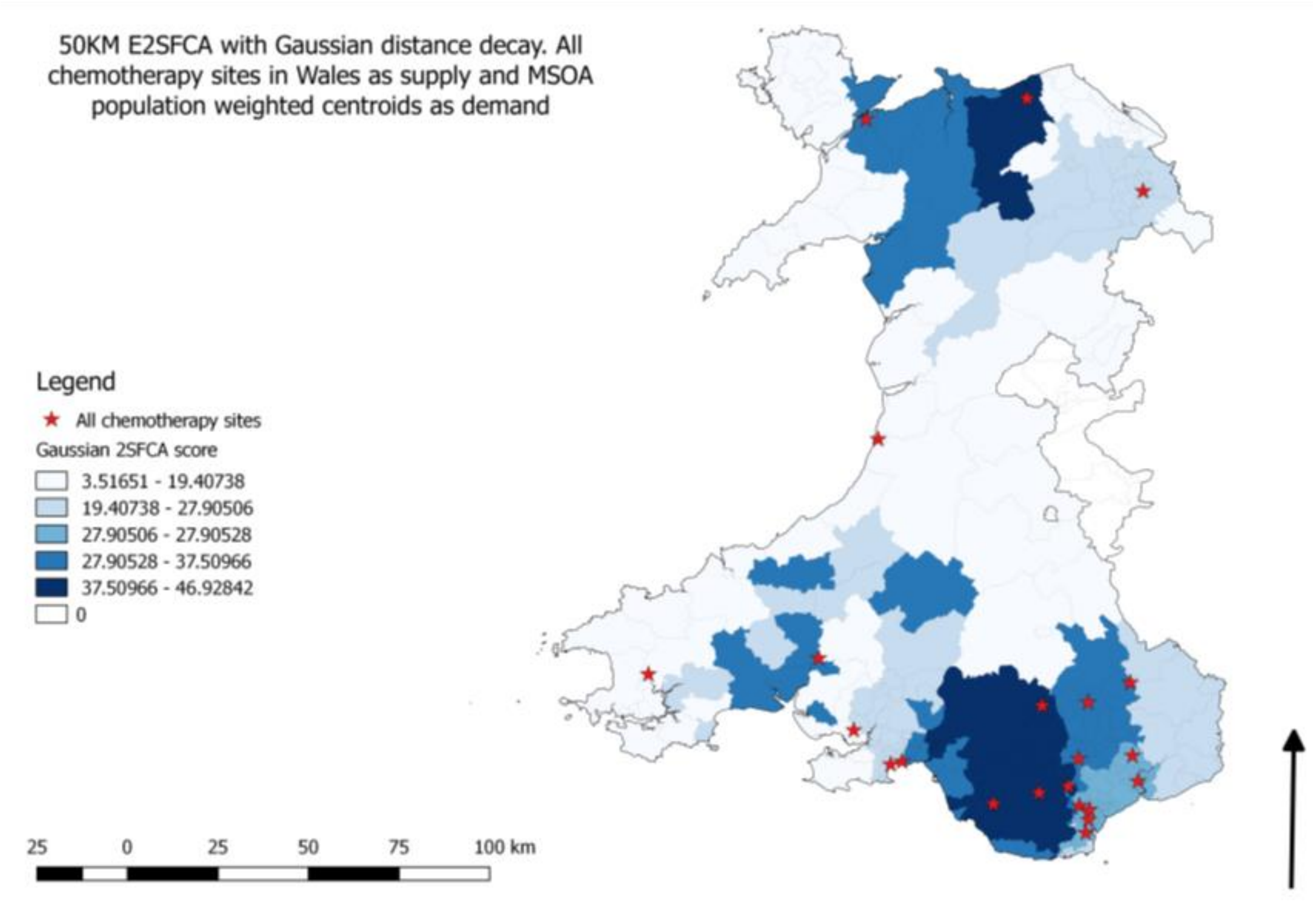


Figure 6-27 50km Gaussian E2SFCA results at MSOA scale

Table 6.7 is very similar to table 6.6 and it indicates that the 35km catchment area already included most of the results. The differences between the data were vast and it would be advisable to work with the data at the smallest scale where available. The results indicated that although they are not as intricate it is still possible to get a good understanding of accessibility with MSOA level data.

50km	**Mean**	**Total results**
OA	1.1	9903
LSOA	5.9	1837
MSOA	27.6	405

Table 6.7 50km means and number of results

Although the preference would be to work with data at the smallest scale it is often difficult to obtain data at a fine resolution especially with healthcare records as it needs to be respectful of individuals' data. These results would indicate that on a country sized study it is possible to gain useful results from MSOA data, although LSOA data would be preferable and enables comparisons to be made with deprivation measures such as the Welsh Index of Multiple Deprivation (WIMD).

6.4.2 Network data

Following on from section 4.5.4 a number of tests have been completed on OSM and OS Open Roads and Google routing data. To understand the differences between the data a comparative test was completed using the two open data sets and Google. 100 postcodes in Wales were randomly selected from a dataset of Tenovus users which represent different areas within Wales. The postcode data was arbitrary and was chosen to test random routes within Wales across the 3 separate networks. Routes were computed to and from each using each network. Routes were of varying size and utilised different types of road. It is anticipated that Google uses a more complex algorithm (traffic variations) to establish routes than the Dijkstra which is utilised by the tool which may make their journeys longer in distance.

Table 6.8 shows that Google times are the highest with OSM next, followed by OS Open Roads. This makes sense as OSM has different road speeds assigned to each road, and OS Open Roads is using the seven different categories which is less accurate. The distances between OSM and OS Open Roads are very similar with a slight difference in the mean distances which highlights the differences in speed limit attribution.

	OS Open Roads (Time)	**Google (Time)**	**OSM Time**	**OS Open Roads (Distance) KM**	**Google (Distance) KM**	**OSM (Distance) KM**
Max	02:58:17	04:14:00	03:32:52	270.8	329	270.5
Min	00:01:12	00:05:00	00:01:31	1.2	1.7	1.2
Mean	01:12:12	01:42:00	01:26:57	110.1	125.6	111.5

Table 6.8 OSM, OS Open Roads and Google comparison

In addition to the statistical data, Figures 6-28 and 6-29 show the pattern across different journeys and suggest that if time is to be used, then OSM provides a time which is closer to that produced by Google Maps.

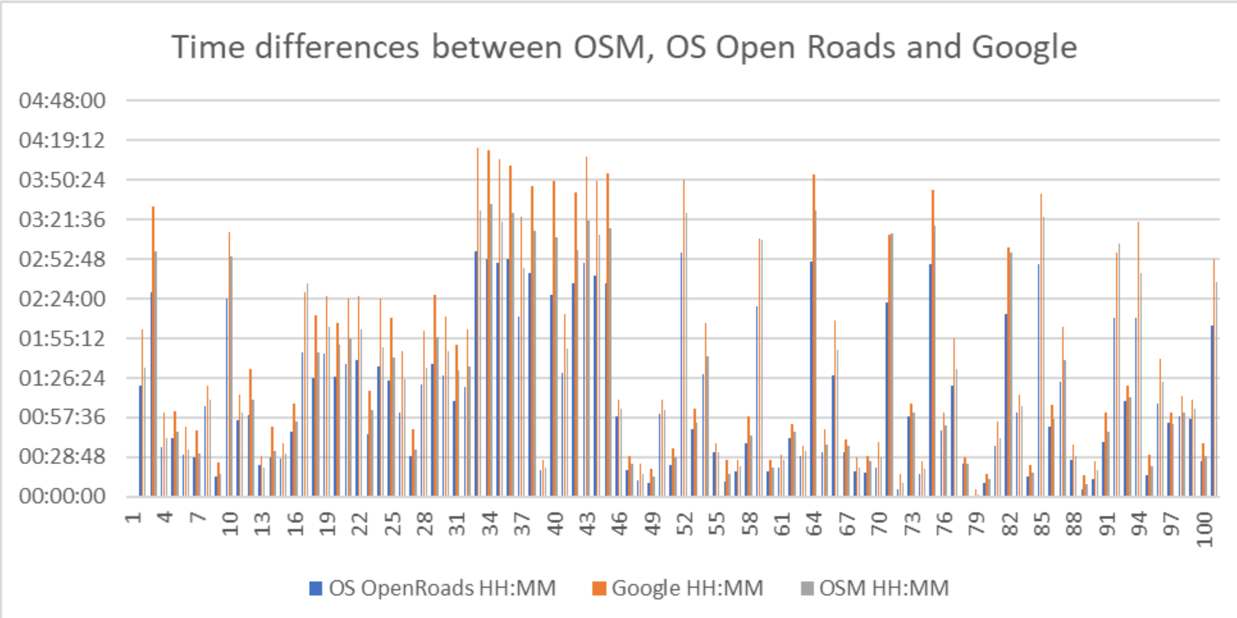


Figure 6-28 Time comparison between Google Maps, OSM and OS Open Roads where the y axis is the time taken to complete the journey and the x axis is the journey number

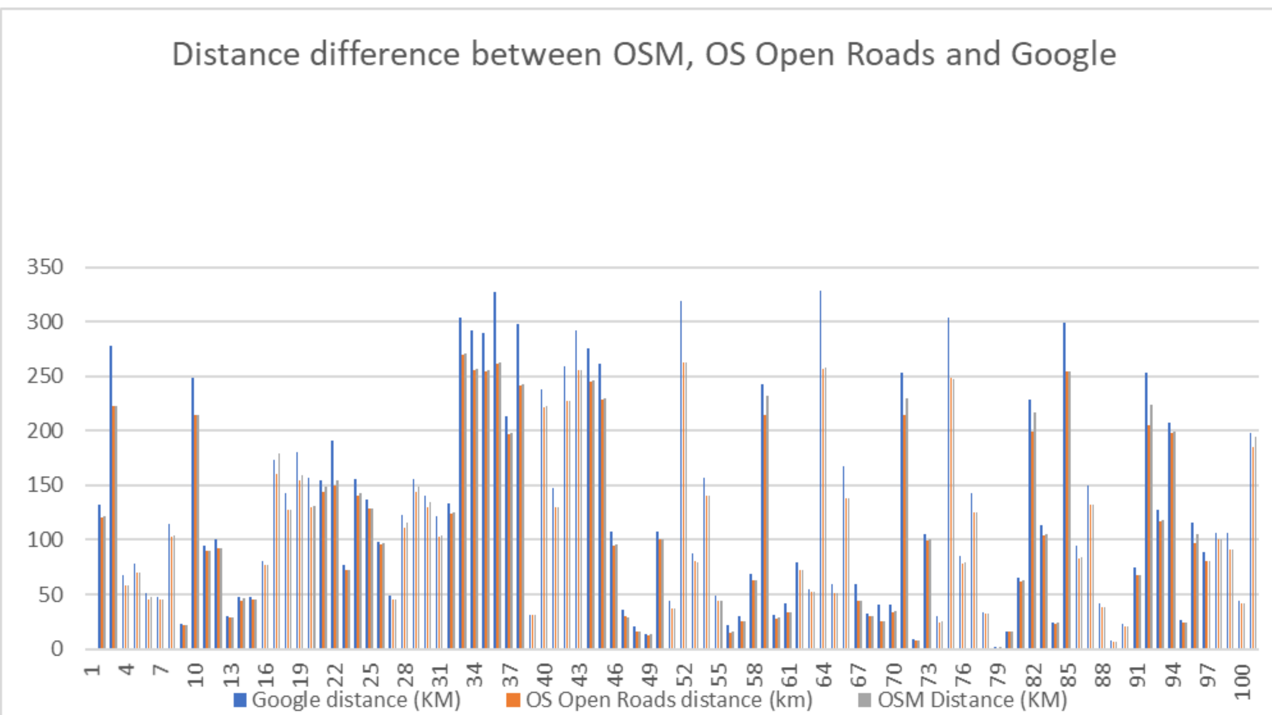


Figure 6-29 Distance comparison between Google Maps, OSM and OS Open Roads where the y axis is the distance taken to complete the journey and the x axis is the journey number

The main variance in distances is between Google and the open source data sets. This may be due to the fact that Google picks the quickest route, and this may not always be the shortest. There are some differences between OSM and OS Open Roads which can be explained by the quality of the OSM data set, and the fact that it more accurately accounts for one-way streets, tunnels and bridges.

The implementation for OSM data is more complicated, it requires the user to change the spatial reference identifier (srid) and import the data using one of the available importers. These may be daunting to a non-expert, and this should be kept in mind when deciding the best data set to use in an operational environment.

Table 6.9 shows the times taken to compute different E2SFCA calculations using the total population of each LSOA as the demand and the number of beds or chairs at each location multiplied by the opening hours as supply. These times reflect the more complex nature of the OSM data set and although the compute time is longer the quality of the results justifies the increase.

Catchment size	OSM	OS Open Roads
10,000	0:51:18	0:43:57
30,000	0:54:72	0:46:56

Table 6.9 Time taken to compute analysis using OSM or OS Open Roads

6.4.3 Network join comparisons

With the knowledge that the OSM dataset is more accurate and only slightly more computationally heavy, it is possible to recommend the OSM dataset wherever possible. The OS Open Roads data set is still capable of providing meaningful results, but is not as accurate and unless road speeds are obtained from an alternative source it is difficult to suggest its use when both are available.

To ensure the best fit algorithm is used, it is important to understand the differences between connecting data points to the network. There are several ways in which this can be done. It is possible to connect the points to the nearest node (junction) on the network, and it is possible to split the nearest line on the network (road) and insert a new node. To add a new node requires either more pre-processing when the data set is added, or another function can be used but this makes it more computationally difficult for the program.

To assess the effects this may have, several tests were completed. This section explores the differences that occurred when connecting a point to the line or to the nearest node and what, if any, effect this had on overall accessibility results. These are small details which can make quite a substantial difference to how the program is constructed and the results it provides. Ideally, every dataset would have each point joined to the network at the appropriate position by an individual going through each one and selecting the position/positions that the datapoint connects with the network. This does not work when using centroids as they are positioned based on population and not networks. For the addition of supply such as hospitals this is possible, but impractical when working on a large scale such as a country.

This leaves two options: to connect the point to the network which involve using the nearest node on the network (Figure 6-33), or the nearest point on the line of a network (Figure 6-32). By utilising different functions, it is possible to do both and the findings presented in this section explore whether it makes any ‘practical difference’ to the results. Within this section the idea of connecting to the line or nearest node is explained and the differences between the two are shown through a series of examples and scenarios.

Figure 6-30 shows a small-scale example of the differences between joining a point to the nearest line on a network or the nearest node. The results are similar and they both show results that can be interpreted visually.

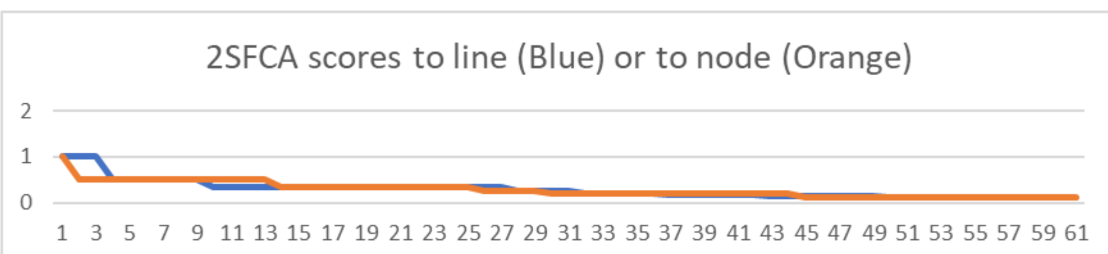


Figure 6-30 Comparison of differences in 2SFCA scores between OSM and OS Open Roads where the y axis is the FCA score and the x axis is the series of results from 1 to 61

The differences in table 6.10 show that even though both are similar connecting to the node has removed 3 results. These removals increase the accessibility in other areas and in general make the results less accurate. This does not necessarily mean that these points are wrong as there is the potential for a site to connect to the nearest line and be closer to the target than its actual entry point to a road, or in the case of hospitals there are potentially many entrances and exits to the road network and it is implausible to get an accurate representation of this.

	To Point	**To Line**
AVG	0.2931	0.2786
Min	0.11	0.11
Max	1	1
Sum	17	17
Number of Scores	57	60

Table 6.10 Comparison between connecting to the nearest point on a line or the nearest node to 4 decimal places

Figure 6-31 shows the difference between connecting to the line or connecting to the nearest node. Node 376562 needs to be attached to the network for the network distance to be recorded, to do this it can be connected to node 376562 or the brown dot which represents its nearest position on the network. There is only a small difference between these two connections and the difference in the output is likely to be negligible. It is also important to clarify that the entrance to the site of node 376562 may occur from any of the different lines that surround it.

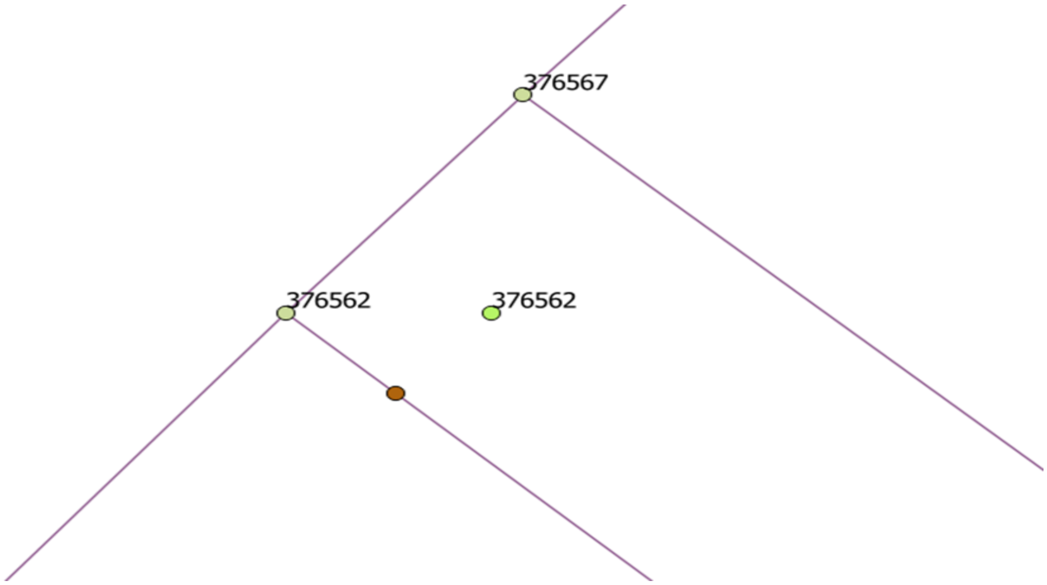


Figure 6-31 Example of nodes on a network where the line is closer than the nearest node

It may also occur that a node is the nearest position on the network as shown in Figure 6-32. Node 19923 needs to connect to node 19923 as the closest line segment is further away. This is less common than the nearest line being closest but does occur. In these scenarios there is no advantage to connecting to the line but the vast majority of connections are to line. When tested 8.02% or 805 from 10,037 connect to the nearest node when using OA data set and this is similar when looking at LSOA data which is 10.31% or 197 from 1910 connections

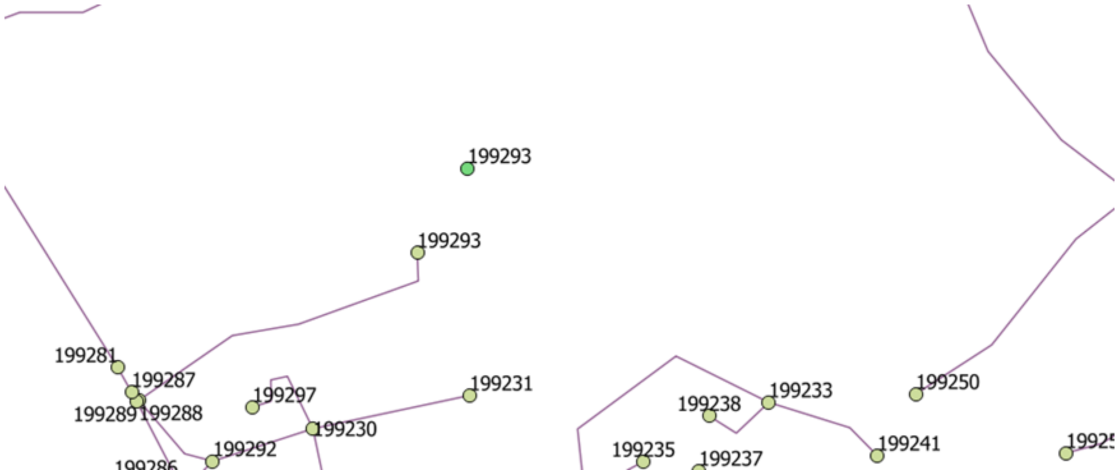


Figure 6-32 Example of nodes on a network where a node is the closest point of connection

This is also represented in Figure 6-33 where node 102446 needs to connect with node 102446. Figure 6-33 shows that although it may be best to connect to node 102446 it is very close to being connected to node 102344 or 102334 which have impacts on the routes created.

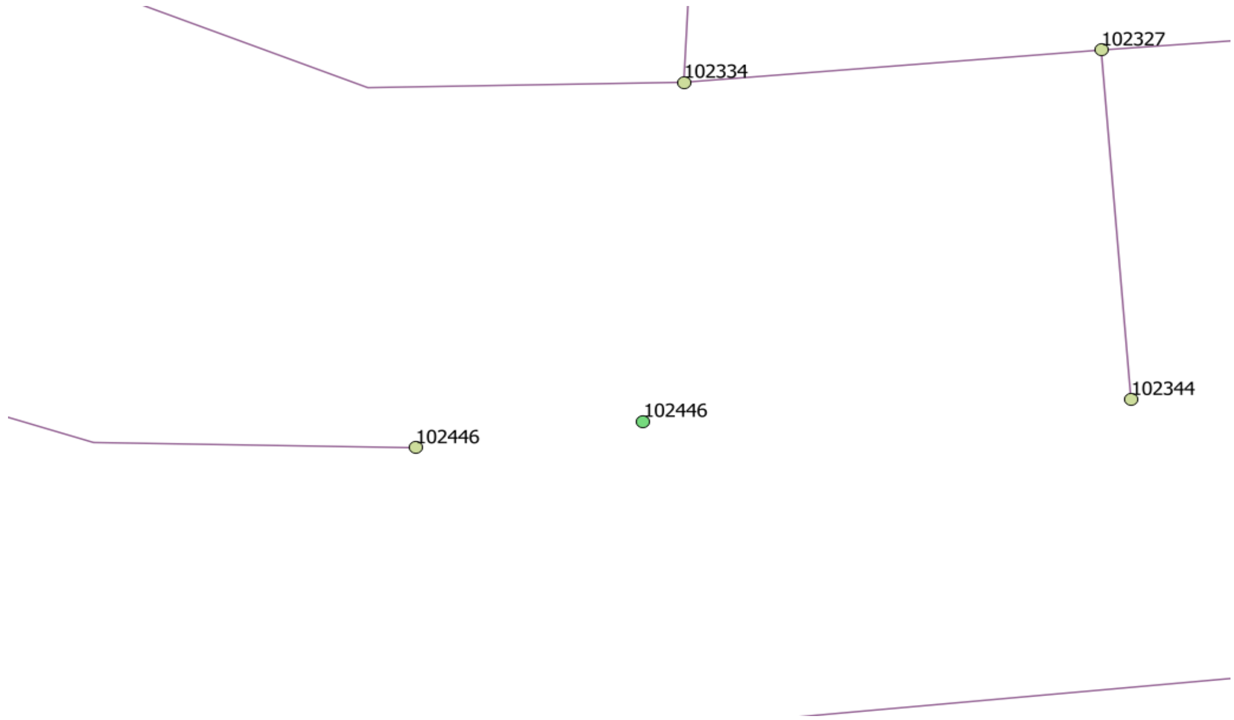


Figure 6-33 Example of nodes on a network where the nearest node may not be the best choice for connection

Figure 6-34 shows a situation where the nearest node is considerably further away than the nearest line. Point 166065 can connect to the network on the line which would be the most reasonable position to join, but by connecting to the nearest node it would connect to node 166065 which adds extra distance or time to the route. Although this is a small distance or time, when these are accounted for across the network, they have the potential to make the results less accurate than they could be. This is a more common occurrence and explains why connecting to the nearest point on the network is the best way to work with routing data.

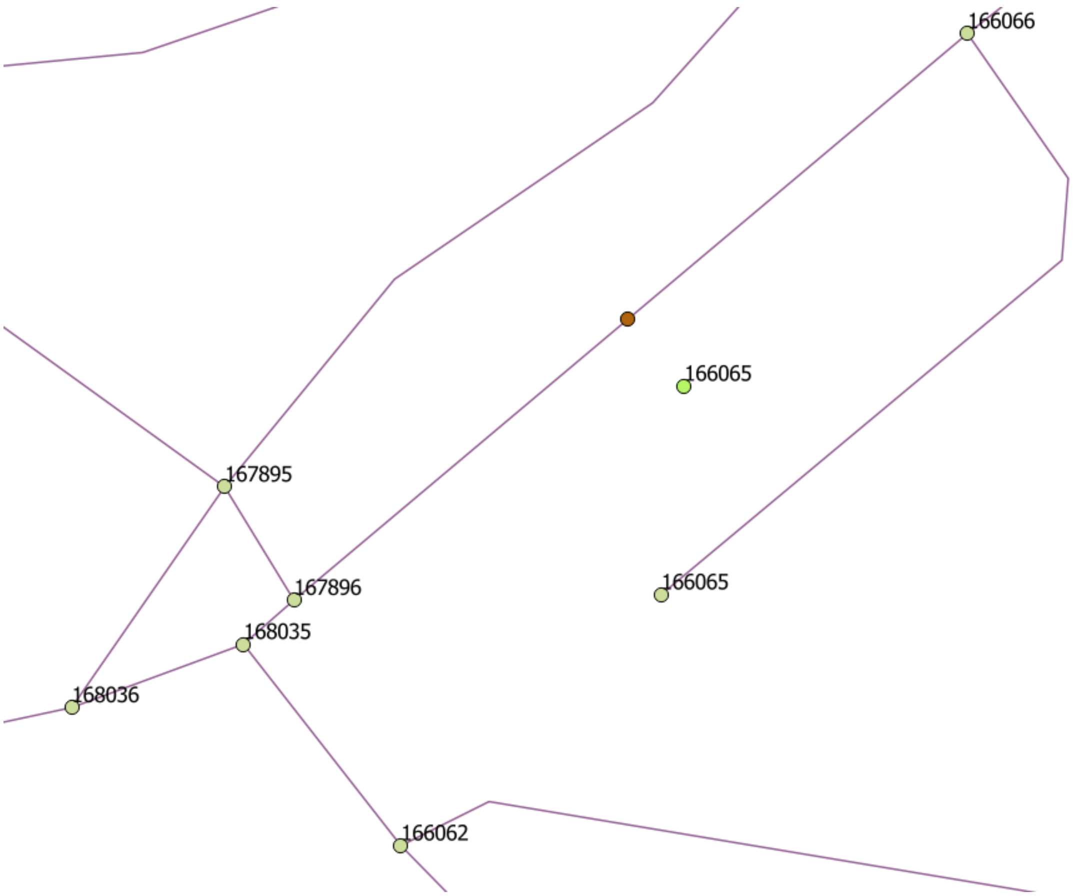


Figure 6-34 Example of the nearest node on a network being considerably farther away than the nearest line

Figures 6-35 and 6-36 show the differences that arise in the output scores when basing the calculation on an algorithm that connects to the nearest node (Figure 6-35) and to the nearest place on the network (Figure 6-36). The maps themselves are very similar and this suggests that when reviewing the results from a practical standpoint there is typically very little difference between the two methods.

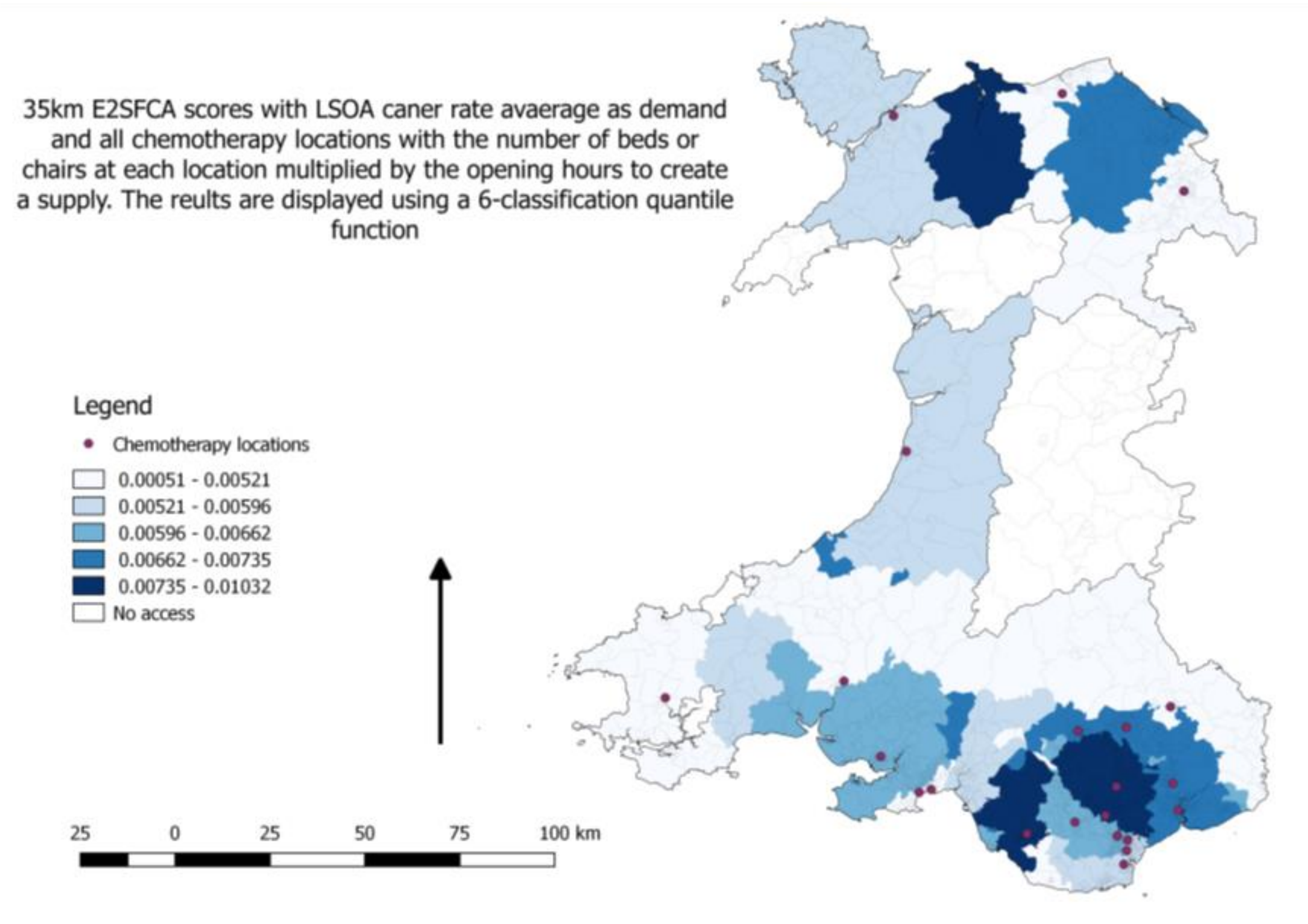


Figure 6-35 E2SFCA score when connecting to the nearest node

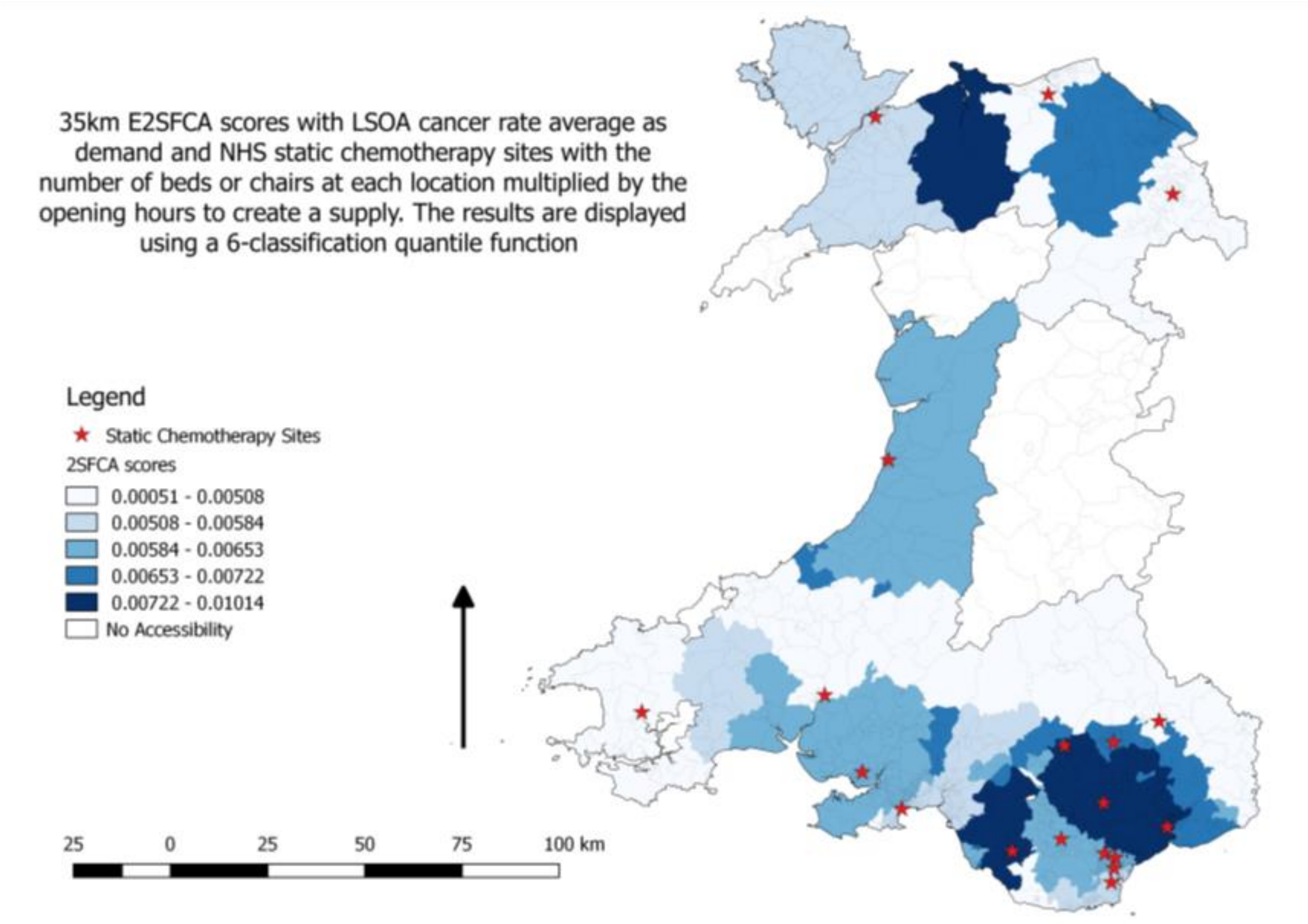


Figure 6-36 E2SFCA scores when connecting to the nearest line on the network

Although neither method is perfect the images and statistics show that it is better for a network to connect at the nearest point rather than the nearest node. That said it is possible to use both of these methods and get interpretable results as there is error in any large-scale network. It is also worth considering that much of the data connected to the network is a single point for a large area, or an aggregated point for population weighted centroids and as such is inherently not perfect to join to a network in any case.

6.5 2SFCA comparisons

Table 6.11 shows the times taken to compute 2SFCA and E2SFCA with a linear and Gaussian distance decay function using the total population of each LSOA as the demand and the number of beds or chairs at each location multiplied by the opening hours as supply. The differences in time taken to compute is very small, and at most practical scales there is very little difference computationally. This indicates that it is better to use one of the more complex methods to measure accessibility as standard on the tool. By using a distance decay function as standard, it allows the non-experts to get a more detailed understanding of their data

Catchment size	**No Distance Decay**	**Linear Distance Decay**	**Gaussian Distance Decay**
10,000	0:51:18	0:50:92	0:51:10
30,000	0:54:72	0:54:38	0:54:36

Table 6.11 Computational time taken to calculate 2SFCA scores with no distance decay, linear distance decay and Gaussian distance decay

Figures 6-37, 6-38 and 6-39 show the differences between 2SFCA and E2SFCA with a linear distance decay and a Gaussian distance decay. The maps show that at a national scale there is very little difference overall, but the differences may be important when analysing this data.

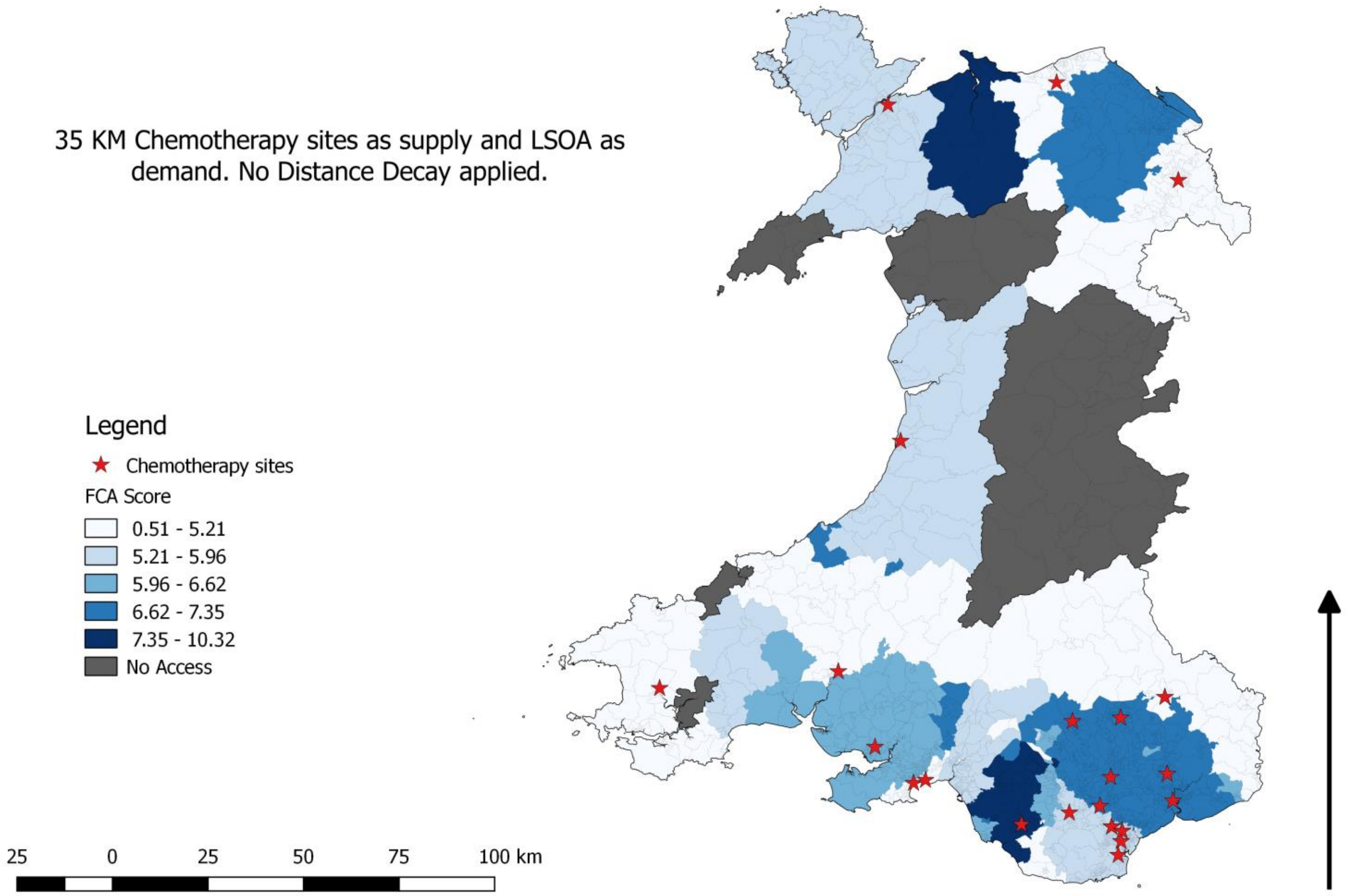


Figure 6-37 Chemotherapy sites as supply, LSOA population weighted centroids as demand at 35km with no distance decay

Both of the distance decay maps offer more detailed descriptions of the data, which is well represented in the middle of Wales where everyone has the same access over a large part of the country. The linear distance decay function addresses this slightly but it is more evident in the Gaussian distance decay map (figure 6-39).

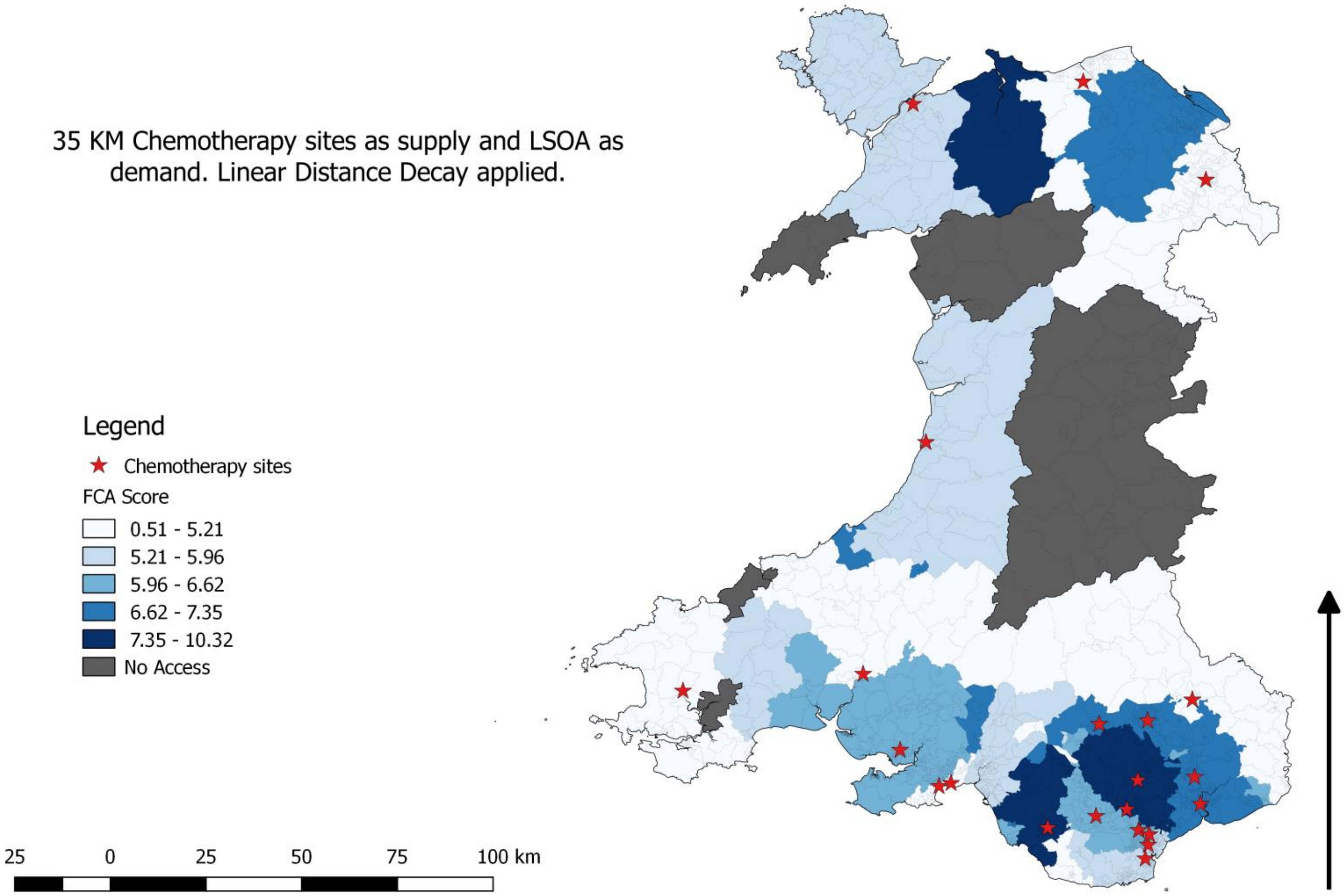


Figure 6-38 Chemotherapy sites as supply, LSOA population weighted centroids as demand at 35km with linear distance decay

The difference throughout the rest of Wales is more subtle but visible between the maps. The north west has a more detailed interpretation with the Gaussian distance decay and the gap in the south east of Wales is more pronounced with the Gaussian function applied.

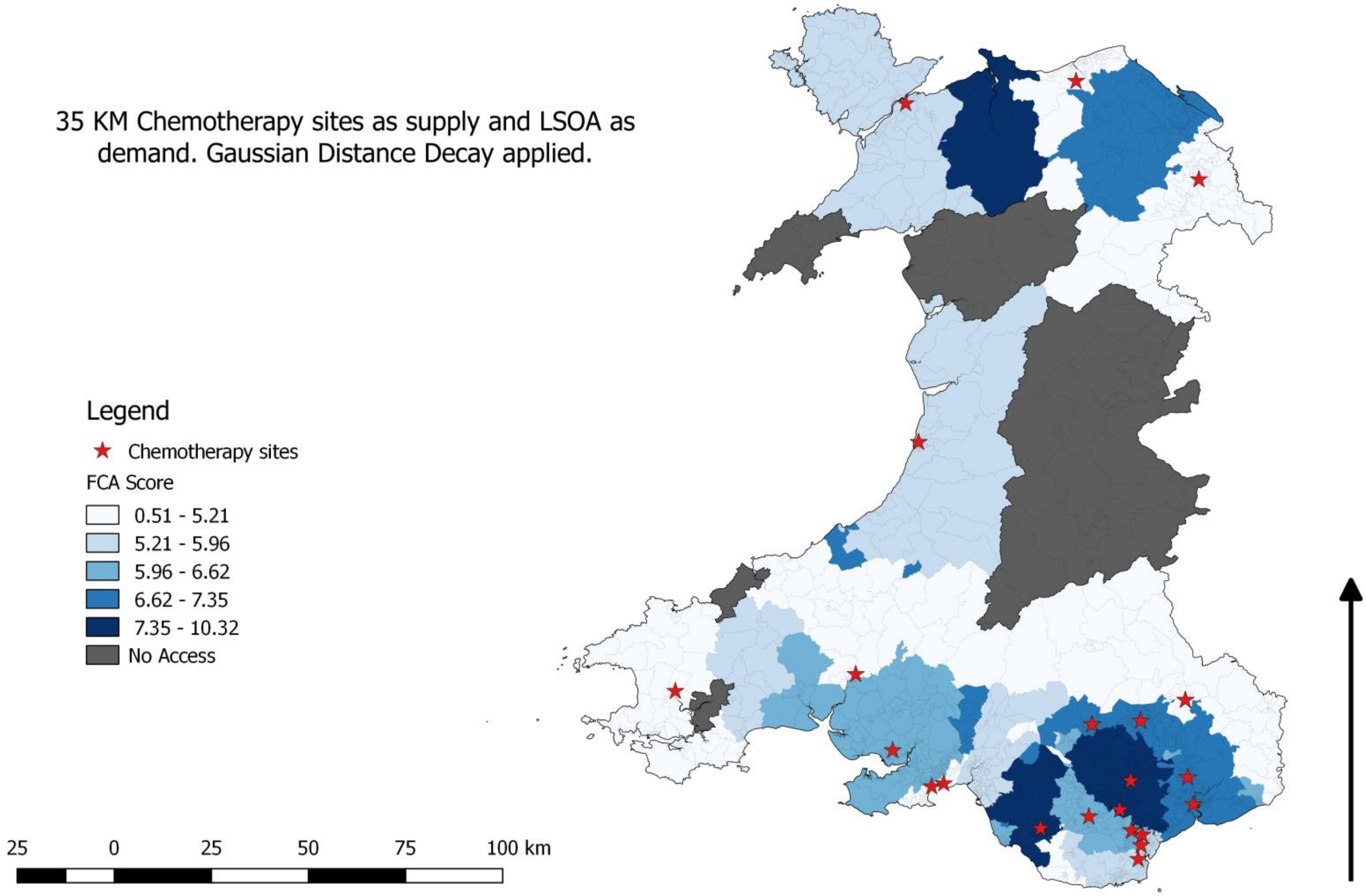


Figure 6-39 Chemotherapy sites as supply, LSOA population weighted centroids as demand at 35km with Gaussian distance decay

Table 6.12 shows that the maximum and minimum values (to 5 decimal places) separate when using distance decay functions, which allows for a more detailed description of the data. Showing that those closer to a supply point have better access than those further away is more reflective of the real situation and produces results that fit more in line with expectations.

	AVG	Max	Min	Sum
No DD	0.00954	0.05714	0.00129	17
Linear DD	0.00954	0.08149	0.00024	17
Gaussian DD	0.00954	0.14449	0.00006	17

Table 6.12 Results comparison between different distance decay functions

As discussed in section 4.3.3 based on the utilisation, data provided by Tenovus is more likely to be Gaussian in nature and this is used wherever possible. There are several issues with using the Gaussian distance decay for non-experts, as it requires additional inputs compared to the linear distance decay. Nevertheless, it should be offered as an option so that someone who understands the inputs has the ability to run the most suitable analysis.

6.6 Chapter Summary

The results section highlights the principal findings from the study. This project has several dimensions which encapsulate different types of findings. Understanding the performance of the tool and stress testing it allowed for comparisons with the other available tools that have been developed using proprietary tools (Higgs et al., 2018; Langford, Higgs and Radcliffe, 2018). This was key in ensuring that the software solution that was developed is capable of working with operational datasets in an acceptable timeframe.

A case study was completed to investigate the effects of moving the Tenovus mobile units to a variety of geographical locations. This case study not only permitted an opportunity to better understand chemotherapy provision within Wales but allowed the tool to be showcased and show its potential. Images created from the case study could be used to highlight areas with poor accessibility, and hopefully feed into plans to increase accessibility to those potential patients living in those areas of Wales that are currently poorly served by this type of healthcare provision. A large number of visualisations were created during the project; these have been highlighted as the output is useful in representing very large datasets. It would be very difficult to understand the accessibility scores without these visualisations, and they have a potential to help those who are not experts within the field to plan services.

Investigating the impact of employing different data sets was a huge part of the project with the ultimate aim of making recommendations regarding the suitability of certain data sources. Findings from the study highlighted the subtle differences in the end results from using data at different scales. The variety of network data available is large, and a comparison between the two popular open data sources (OSM, OS Open Roads) and Google was completed to understand the differences in the data. This comparison highlighted the advantages of OSM and although it is more complex to create a network, the results are much more in line with the proprietary options available.

As with many parts of this project, there are many ways in which the data could be connected to a network, and two potential options were explored in detail. One had the potential to be more accurate whilst the other offered easier implementation and smaller computation effort which would affect the speed of the program. The differences in these results are of particular interest as a considerable amount of the data they are connecting to the network is aggregated, and therefore itself not wholly accurate. The results suggested that it is better to connect a point to the nearest part of the network, but in practical application it appears to make very little difference.

The shortfalls of the different types of 2SFCA were discussed in the literature review in detail, but many different iterations of the method exist and they attempt to make the method more accurate or more suited to specific data types. This study compared the standard 2SFCA, a linear and a Gaussian distance decay to observe the differences in visualisation. There are obvious benefits in terms of computation and simplicity by using less involved methods and understanding the differences for the end user help to guide the implementation of different methods. The results suggested that the Gaussian distance decay produces the most detailed images, although the linear distance decay also had the advantage of the standard 2SFCA method.

The results provided a different perspective from that normally found when including the use of SFCA methods in measuring access to services. Although they showed that more complex methods provided the most complete results, findings from the use of different spatial aggregations also highlights that at each scale of analysis the methods offer value. The investigations into the data provided additional insight into the quality of the open data networks available and the best options for implementation. The data results highlighted interesting observations and as highlighted in the remaining chapters could be included in future iterations of the software.

Chapter 7 Discussion and evaluation

7.1 Introduction

In previous chapters, the results from applying floating catchment area tools for locating potential sites for cancer services and for monitoring the implications of changes to the overall infrastructure were demonstrated through map and tabular outputs. Chapter 2 illustrated how useful FCA approaches can be in planning cancer services in relation to potential demand and examined the use of such tools in modelling the impacts of different planning scenarios associated with siting of cancer services. By drawing on case studies, the advantages of such approaches over the use of 'traditional' accessibility measures in siting facilities, and the use of open sources of data to help run these models within over source GIS frameworks was demonstrated. Chapter 4 highlighted design considerations and the development of such tools for non-GIS experts and drew attention to a wide range of usability issues that needed to be addressed in order that planners and policy makers have a suite of open source tools that can be used to plan the location of mobile services. These sizeable chapters provided a body of empirical evidence that showed how such approaches can benefit healthcare professionals concerned with providing locations for cancer support, screening and chemotherapy services.

This chapter aims to present the key results from this research, and compare these findings to those of previous research studies concerned with examining the use of accessibility measures in general, and floating catchment areas techniques in particular, in health domains. In so doing, the aim is to document the principal advantages of an open source approach to investigating access to services/facilities, and consider the main themes that have arisen from this research. By highlighting the potential of such tools to address real world planning issues such as the location of cancer services, and demonstrating how such tools can be packaged in user-friendly formats to enable those charged with managing such services to examine the implications of changes to the location and characteristics of facilities, this research has advanced previous studies in a number of ways – these are discussed further in section 7.3. During the course of this research, a number of potential limitations of the approach have been highlighted and these are discussed in section 7.5 in terms of the potential impact on research findings. Finally, the policy significance of this research is discussed in relation to the overall planning of cancer services in Wales and the relevance of such work to wider national and international contexts.

7.2 Summary of the Research Design

This project spanned across several areas of speciality, computing, human geography and usability. The ability to accurately address these areas was key in the success of the project. The work initially focused on the different research of methods that would be suitable for a project such as this (Chapter 2). This base level of understanding of the current research assisted in the design of the tool, and allowed a proof-of-concept tool (Chapter 3) to be developed. The proof-of-concept tool used a very popular GIS to compute the FCA calculations using straight line distance. This tool assisted in focusing the requirements of the project and allowed for time to investigate the different options available for development.

Usability was a very important part of this project, and as such many models were researched to ensure a positive outcome (Chapter 5). One of the similarities amongst many of the models was to keep the user at the centre of the design, and to do this there was a high level of contact with Tenovus at each stage. Initially high-level conversations took place with the key stakeholders to understand the requirements of the organisation and to help them understand the potential of these accessibility measures. Once the requirements had been discussed it was important to show the initial results and gain feedback in user study 1 (section 5.5.1) this was followed with a further user study (section 5.5.2) which was able to provide feedback and ensure that the tool was usable by its intended users.

There are numerous ways in which a tool like this could be developed. It has been integral to the project to investigate these options and provide feedback on many of the free and open source development options available (section 4.2.1). There are a number of different approaches that could be taken to develop a tool such as this and the research has been designed to investigate the effects of these choices relating to how a network is formed or the different ways in which the FCA methods can be implemented. The goal was to see what, if any, effect these changes had with real data at a national level. This knowledge should be able to assist researchers and industry to understand the effects of scale, or the nature of algorithm used to investigate spatial patterns in accessibility.

Real data was key to understanding the current state of play in chemotherapy provision in Wales. Gathering and presenting this data allowed for the project to have real benefit to key stakeholders in Wales who plan and deliver cancer services. The data gathered by this project (section 4.3) gives an accurate insight into the supply of chemotherapy and lymphoedema treatment, as well as the demand for these services. Understanding the impact of variations in demand on service provision was important and the data provided by Tenovus on utilisation of their services provided an important insight into the way people interact with treatment, which can be used by others.

A case study was completed to assist Tenovus in understanding their current impact upon chemotherapy services within Wales as well as to gain further knowledge on how adding additional resource to areas would affect the chemotherapy provision as a whole. This case study used the data that had been gathered and produced visualisations which could be used by key stakeholders to plan services and monitor the impact of service reconfiguration.

7.3 Key findings

During this project there were several key findings, these can be summarised as;

- Creation of a tool which can be used by non-GIS experts to return complex accessibility scores.
- Visualisations which provide an overview of the current chemotherapy provision within Wales (at a national level, including non-NHS parties) including accurate demand side travel data.
- A network comparison between OSM and OS Open Roads.
- Algorithm comparisons (network and distance decay).
- A case study using live data to show areas which could enhance accessibility to chemotherapy throughout Wales.
- Data gathering of supply and demand side metrics.
- FOSS GIS comparison of the main solutions available.
- Adding to previous literature with accurate demand side data (Luo, 2014, Dai, 2011) and national scale cancer service provision (Shi et al., 2012).
- Conducting full user experience studies with people involved in cancer service provision on FOSS GIS.
- Addressing multiple policy requirements.

The tool created for this project can be used by non-experts to create complex accessibility maps. As this is all completely free and open source it allows anyone who has an interest in healthcare accessibility to use these advanced methods. The tool has not reinvented the methods used but has tried to package them in an easy-to-use format which can potentially have a real-world impact. The way the program is written should enable flexibility for those who want to use and adapt it further by utilising a series of SQL commands which could easily be used as a standalone series of inputs, or with a different programming language to enhance a current GIS application or create new functionality. By making these methods more accessible it is hoped that further enhancements could be made to ensure wider utilisation of the methods.

Maps created during this process were used to assist in the development of the tool and also to help Tenovus understand the current state of chemotherapy provision within Wales, as well as several of their other services. The visualisations take complex accessibility scores and display them in an easy to interpret way. These outputs assisted in the planning of future unit locations, and also provided a tool for them to show the effects of their services. Conducting a scale comparison assisted in showing that although the smaller the scale the more detailed the output, there is a scale which describes the data well enough for most purposes. In this example the LSOA visualisations were completely acceptable for maps at this scale and although a small amount of detail was lost, it is still worth completing the analysis at this level as the data is often unavailable at smaller scales. It is possible to dis-aggregate data to smaller scales but this would also not have an ideal result as data aggregation does not provide a more realistic result.

The network comparison was an integral part of the project and the results of the comparison helped with the general understanding of those open data network datasets that are available to researchers. Using Google maps as a standard and investigating the differences in results in terms of times per route and distance per route showed the quality of the OSM road network and the limitations of the OS Open Roads network data set. The differences in implementation are large and often the network database used within these models depended on the prior experience of users and the overall aims of the project.

The development of appropriate documentation and user-friendly guides was a large part of this project, and understanding how users interact with a system helped in creating detailed and well-thought-out documentation to ensure that the user can follow each stage of the floating catchment area methodology. The development of the UI incorporated many prompts which should limit the need for documentation, as after the first one or two interactions with the program the on-screen help should provide adequate assistance.

The case study helped to provide meaningful results to Tenovus and assisted in showing the capabilities of the tool. Using a free to access dataset it was possible to find those supermarket carparks available in the UK that could host the mobile units in order to identify those locations that could have an impact on the current chemotherapy provision, and then visualise the results of these changes so that potential partners and stakeholders could make a judgement on the efficacy of future plans. It was also possible to understand how much provision a place needed in terms of days at each location by increasing the supply points for the selected units. This case study was reasonably small in scale but fulfilled the purpose of showing the potential of the tool.

A large part of the project was spent investigating the benefits and negatives of connecting datapoints to the road network. There were several ways in which this could be done and all have their benefits. It was easier to connect to the nearest node on a network computationally because the network does not require the lines of the graph to be split and a new node to be inserted. Although this was easier computationally, it was not as accurate as connecting the datapoint to the nearest line on the network which required the network to split a line and insert a new node. This was also not 100% accurate as the nearest line may be a road with no access point. To get this type of accuracy would have required going through each individual point and investigating Google maps' street view or similar, which was impractical. The fact that the datapoints are often aggregations and are therefore not an actual place meant that the data was itself not particularly accurate. The question this tried to address was; is it worth connecting to the line when the data being connected is aggregated, and would this make any difference to the results? The comparison showed connecting to the line appeared a reasonable assumption to make and that there were several examples where the nearest node was a considerable distance away which could have skewed the results. Consequently, the to-line methods have been implemented but the to-node code is still available in the program.

Using informed demand side metrics helped the project to have some real impact. Using historical Tenovus data enabled a genuine insight into those utilising chemotherapy services within Wales and as a result could be used as a guide for setting the parameters in the E2SFCA methods. The fact that the data-decay utilisation suggested that the tailoring with distance was Gaussian in nature and had a clear cut off at around 35 kms helped inform the actual implementation of the models. The data can also be used for greater insight into the users of such facilities to focus the demand side variables either by sex or age.

Completing a FOSS GIS comparison increased knowledge within this space. There are many FOSS GIS solutions and covering all of them would have been very difficult, but by comparing the main options available and explaining their network analyst capabilities allow future research in this area to be completed more smoothly with less groundwork required. One difficulty with this type of comparison is that the speed of development on each platform is fast, and keeping abreast of all of the updates is challenging. The main findings of this project highlight how this tool can be of use when evaluating and planning cancer provisions. The project has presented an up-to-date overview of the current chemotherapy services within Wales; this knowledge helps healthcare professionals understand the current state of provision and highlights potential gaps in cancer services across Wales (see section 6.2).

Many of the studies conducted in the peer reviewed research (section 2.7) are concerned with cancer screening or very large-scale studies investigating cancer centres in the USA. This study built on the work conducted within those studies in several ways, firstly it provided a detailed understanding of chemotherapy provisions in Wales with accurate supply and demand metrics. Most studies do not have an accurate estimation of the extent of catchment area to include within FCA models as they are unable to estimate the distance travelled by potential users. Understanding and including accurate behaviours of the demand variable increased the validity of the results. Secondly this project moved from researchers being able to provide results to the supply organisations themselves. Allowing these studies to happen at much more frequent intervals provided more data to the providers, and hopefully assisted in real time assessment of the provision.

Tools such as this can be used in a variety of ways to assist in decisions relating to the planning and provision of services. They can be used to evaluate the current provisions in place. This allows for greater knowledge for the key stakeholders, and any gaps that appear in the service provision can be evaluated and plans made to address this shortage. The methodology used in this project has highlighted that even with relatively high-level data it is possible to gain a more realistic perspective of accessibility by using these methods in tools such as this.

7.3.1 Evaluation of research objectives

The different aims and objectives in this study can be appraised by collating the results gathered during the project and evaluating their usefulness for organisations such as Tenovus in their operational and strategic planning of service provision, and the contribution the work has provided to the different areas of focus.

7.3.1.1 Research Objective 1

The objective to examine the utility of accessibility models based around the use of floating catchment area (FCA) methods in healthcare studies with specific emphasis on the provision of cancer services has been met with in several chapters of this thesis. Such an evaluation includes an examination of the utility of accessibility models based around the use of floating catchment area (FCA) methods in relation to previous studies concerned with examining the provision of cancer services. This builds on the findings of those studies highlighted in Chapter 2 that have used FCA approaches to investigate cancer service provision, the work conducted in this thesis has expanded on the study by Shi et al., (2012) which investigated access to cancer services across the USA. By evaluating access to chemotherapy services in Wales this project is on a smaller geographical scale but agreed with some of the findings; such as the different distance decay gave more detail but did not have a large effect on the overall pattern. In section 6.5 it was clear that although the most detailed results were present in the Gaussian E2SFCA, the linear E2SFCA showed a similar pattern across Wales.

7.3.1.2 Research Objective 2

The objective to investigate and evaluate the development of solutions based on the free and open source ethos that can be utilised to help administrators and managers and other non-GIS experts to plan the provision of healthcare (cancer) services was achieved during the analysis contained in Chapters 3, 4 and 6. There was a vast array of options which could have been investigated; by investigating the different offerings with a focus on how they could be used for healthcare accessibility studies such approaches provided an alternative perspective to the existing studies that could be compared to the use of alternative open source solutions based around the use of Grass or gvSIG.

7.3.1.3 Research Objective 3

The objective to illustrate the practical application of these tools by using data on the current availability of cancer care services in Wales using implementations based on FCA solutions was met in chapter 6. By gaining access to current static site data, it was possible to produce accurate visualisations of the current chemotherapy provisions within Wales. These provided an insight into chemotherapy provisions which were not publically available, and could be used to assist in the planning of future chemotherapy provision. The data is only for Wales, and as there is a large border with England the visualisations may not show the most accurate picture they did not address cross border movement.

7.3.1.4 Research Objective 4

This objective to collate and analyse a number of datasets which can be used to give a meaningful insight into the current provision of chemotherapy services as they exist within Wales was addressed in sections 4.3 and 6.4. This proved to be a challenging objective given the paucity of data in the public domain and variable quality of some of the data sources. This highlighted some of the lessons that need to be learnt if these methods are to be widely employed to examine accessibility to healthcare provision. For example, most of the health boards report their services differently and finding common metrics at a suitable scale was impossible with the data that is publicly available. After an FOI request it has been possible to collate data on some aspects of provision and provide some preliminary estimates of service availability and characteristics. With many FCA studies it is very difficult to understand the behaviour of the demand and to having access to data which gave the journeys taken by each Tenovus mobile unit allowed for more accurate catchment areas to be used and also assisted in implementing the most suitable distance decay function to get the best possible results. The studies by Luo (2014) which uses a continuous distance decay, and the study from Dai (2011) which uses a Gaussian distance decay have been enhanced by incorporating this data. This study used similar methods but used accurate demand data to scale the catchment areas, this demand side travel data is very rare and being able to implement it with these advanced measures provided an uncommon insight.

7.3.1.5 Research Objective 5

The objective to demonstrate that the use of GIS-based accessibility models to examine the supply of mobile services in relation to potential demand for cancer services was met in chapter 6, which showed how the use of these models highlighted variations in accessibility within Wales. Many different variations were utilised such as different distance decay functions (linear, Gaussian, none), different scales (OA, LSOA, MSOA), different demand data (chemotherapy, lymphoedema, age) and different supply data (static site, all sites, mobile unit data). Although much work was done on this objective there is scope to go further and investigate different methods and models and compare them. This would have been a useful addition to the methodology adopted by demonstrating some the key differences between some of the less advanced methods and the FCA methods used here which have been the focus of many of the earlier studies on variations in healthcare provision reviewed in Chapter 2.

7.3.1.6 Research Objective 6

The objective to develop user-friendly software to enable non-specialists to explore the implications of alternative models of provision of mobile services was met in Chapter 5 using human computer interaction models and methods, the ISO 9241-210 model was selected as a basic framework because it ensures the user stays at the centre of the development and design process whilst working in an iterative pattern where a design is tested and re-evaluated. This study followed work from Preece, Rogers and Sharp (2004) which suggested the use of multiple data gathering techniques to mitigate the risk of missed behaviours or data and the discount usability testing (Nielsen and Landauer, 1993) to ensure the best cost-benefit was achieved, using this technique allowed for maximum return of data for the time invested by the people at Tenovus. The inherent problem in creating a complex program to be used by non-experts was that there needed to be someone technical to assist in the set-up and updating of data. This was been managed by creating detailed documentation tailored to particular user responsibilities.

7.3.1.7 Research Objective 7

The objective to test the usability of these tools with professionals concerned with helping to deliver cancer services has been met in the user studies conducted (section 5.5) these studies used participants whose job was to assist in the planning and provision of cancer services throughout Wales. The user study participants were from a wide range of backgrounds within this sphere from administrators to management and provided a unique insight. The results show the IT literacy of people within the field and their capability to learn new techniques and skills. This objective has been met and provides interesting insights into the type of typical users positioned within a charity concerned with cancer services provision.

7.3.1.8 Research Objective 8

The objective to draw on these findings to show how such programs can be used by the profession to plan and manage the provision of a wider range of healthcare services has been met in section 6.2, which shows a case study of how these methods have been utilised to provide 3 potential locations which could increase the accessibility of services throughout Wales. By utilising data of large supermarkets within Wales it has been possible to postulate several positions which could have an impact on the accessibility to chemotherapy throughout the Country. The case study shows that it is possible to provide results which can assist in improving knowledge of accessibility and has the potential be used as a guide for key stakeholders.

7.3.2 The application of floating catchment area tools in planning service locations

FCA was shown to be effective for healthcare accessibility (chapter 2) and specifically within cancer studies (section 2.8). This empirical evidence showed that FCA could be used in projects such as this and that it was able to produce insightful results. In large scale cancer studies such as Xu et al., (2017) FCA methods were shown to give a more complete picture of accessibility to cancer services, and Donohoe et al., (2016) showed that smaller scale studies benefitted from the FCA methods when investigating the links between screening services and cancer rates in Alberta, Canada.

There were many methods available to measure accessibility (see section 2.2) and the FCA method had an advantage over many of the other choices. Apparicio et al., (2008) suggested 5 commonly used methods and the non-gravity-based models all accounted for either the supply or the demand but did not account for both. A major advantage of the FCA method was that it measured the amount of supply available and also the potential demand that had access to it. Another advantage of the FCA method was the amount of research that had occurred for healthcare specific accessibility, and the number of optimisations which had been tested and worked in different scenarios. The gravity model was shown to be useful in measuring accessibility but was more difficult to compute and less intuitive. For a study such as this the FCA methods offered a high level of detail within an acceptable tolerance for computing power and usability.

Currently Tenovus decide the location of their mobile units with collaboration from the participating health board, and approach different sites which may allow them to use space in their car park at certain times. The addition of this tool to the process could affect the impact that the mobile units have on the surrounding population. The ability to understand the effects of static provision on their services provides a great deal more information than they currently have access to, and this should lead to more effective planning of services.

The tool has been designed with the user at the centre of the process with the intention that it is utilised by non-experts, and this capability for non-experts to conduct the analysis allows for a much wider uptake within the organisation. By enabling those whose role it is to plan for chemotherapy provision, the availability of tools to interactively analyse various scenarios can empower stakeholders to make more informed decisions. There were some technical challenges when implementing the software in Tenovus, and these need to be discussed by the internal stakeholders as there are several ways in which it can be used. It could be set up on a standalone laptop which is managed by the super user and is accessible when needed by relevant departments, or it could be implemented as a server with the software installed on the individuals' work machines. Both solutions are viable but as it is used irregularly it may be easier to have it as a stand-alone solution which can be used by whoever needs it at the time.

The proposed solution shown in this study is one which uses a spatial database to store and manipulate the data, a standalone C# program which manages the inputs, and QGIS which visualises the results. By using a spatial database, a number of advantages are available like the security of the data and the speed of the manipulation. It is very difficult to manipulate large datasets efficiently and PostgreSQL/GIS is a very well-maintained solution which is used by many organisations worldwide. The standalone tool was created in C# as this was the expertise which Tenovus had in house and the idea was that it could be further customised at a later date. Alternative user interfaces could be developed using the SQL code, or if someone was familiar with the program, they could set it as a stored procedure or a list of commands, but this limits the scope of usability. The results could be visualised using any program connecting with PostGIS as the table is created and needs minimal input to be visualised. QGIS was chosen for this as it provides a lot of additional functionality that Tenovus suggested would be useful in other parts of their work

7.3.3 Design considerations for user-friendly tools to implement accessibility models

Multiple lessons have been learned from the usability studies that were completed for this project. These lessons are important to note as there is very little usability research focusing on free and open source software, or geographical information systems. It became clear that most of the users were able to follow a process and could complete the tasks with each iteration of the tool, most people now use a computer on a daily basis, and are able to follow basic instruction.

The main areas of interest came from the quality of the documentation which needed to not only explain the different steps in each task, but to explain why they are doing these steps. The documentation needed to have clear visual cues to ensure that if they got lost, they could find their way back. Once the quality of the documentation got better the users found the tasks much easier to complete, and often remembered how to complete the task without the use of it.

The language used was an important factor, and as a researcher it is easy to forget that many of the specialist terms used in computing or geography are not commonly understood. By using clear, concise and simple language it was possible to make the user feel more comfortable, and stopped the need for the user to either carry on with a task not knowing what something meant or feeling embarrassed by asking. Inclusive language in tools such as this is key to the success of the tool, and can play a large part in the utilisation of such solutions in the longer term.

The way a piece of software looks can play an important part in how it is perceived. By mimicking current software in terms of help and file locations it can make a big difference to the end user when they consider their comfort using such programs. It is tempting as a designer to attempt to develop something unique or special, but for usability it is better to make a user interface which is as familiar as possible.

7.4 Challenges of working with Tenovus Cancer Care

The experience of working with the partner organisation, Tenovus Cancer Care, was predominantly good. They were supportive and always interested to listen to ideas and input into the project. Over the course of the PhD one of the noted challenges was staff turnover at the organisation, this made planning more difficult as all of the contacts I had left on more than one occasion. For instance, the IT department left after the first year, the data team changed in the second year and the director of the department changed twice during the project. Those who remained at Tenovus were great at helping to re direct us but it was difficult to keep enthusiasm high within the company without a single point of contact.

During the course of the PhD Tenovus also decided to change how they gathered, stored and manipulated data which proved difficult for the project as the new system was being delivered over several years and was managed by multiple departments. A final challenge was gathering data for the project and finding someone to take ownership for this within the organisation.

Although they provided some rare insights with the data they provided FOI requests needed to be made to obtain further information and the contacts of the organisation could not be leveraged which is disappointing. This said, there are many challenges to overcome in all organisations and the experience of working with those within Tenovus was positive and encouraging.

7.5 Strengths of the approach taken

There are many strengths to the approach taken in this project, and by utilising a multi-disciplinary strategy to focus on usability, advanced geographical methods and free and open source software, the project was able to provide a tool which delivers insight into the current chemotherapy provisions within Wales. There are also advantages to working so closely with Tenovus and gathering real demand data coupled with access to potential users for the user studies. This project was able to draw on an overview of utilisation patterns for those patients who use Tenovus' facilities to inform the choice of parameters within the accessibility models. This data is important in how it has assisted the methods implemented within the tool, as well as guiding how the accessibility studies are set up to provide more realistic results. This knowledge has the potential to assist other researchers when designing accessibility studies as this data is not commonly available. One of the main strengths of this project was having such a close relationship with Tenovus which allowed for a considerable amount of collaboration on the project. It was possible to meet with the key stakeholders in Tenovus and understand what they did and were aiming to do (section 1.4). The ability to conduct multiple user studies at the Tenovus office and have such a large amount of time from them was central to this project. Having considerable time with the organisation over several years allowed the project to keep the potential user at the centre of the process and informed many decisions made.

The user studies conducted at Tenovus not only allowed information to flow from Tenovus, but also enabled key members of staff to be part of the development process and gain a detailed understanding of what this technology can do for their service planning and delivery. Getting this knowledge into the organisation has created a better dialogue and allowed for a more focused tool which can really assist them in their day-to-day duties.

By interacting with Tenovus it was possible to design a case study which has value. The case study highlights the current state of chemotherapy provision within Wales whilst investigating the potential opportunities to add additional resource to areas which are underserved by the current provisions. Many of the choices made during this project were based on the requirements of Tenovus and this should assist other charities and organisations who are in a similar situation to utilise the tool.

7.6 Chapter Summary

In this chapter the major parts of the project have been discussed. This has provided a more detailed discussion on a number of the interesting observations presented in earlier chapters. The application of FCA models have been shown to have real potential to address real world issues such as the planning of cancer services. The case study scenarios have illustrated the types of benefits that could stem from the use of these accessibility measures in the health domain. However, a number of technical and organisational challenges remain to be addressed before the use of these tools becomes mainstream in healthcare applications. To date, there remain relatively few examples of the transferability of these models from their development and implementation in academic studies to real world applications. Therefore, studies of the usability of such tools within organisations such as Tenovus would appear to be particularly timely.

Chapter 8 Conclusions

8.1 Summary

This project involved the use of a range of methodologies drawing on themes in geography, computing and usability studies and the research added to the knowledge in a number of ways. The key conclusions from the project show that there are many ways in which it is possible to build a tool to provide advanced accessibility methods and that none of these are a clear 'best' solution. This project shows the importance of usability studies especially within the FOSS sphere where it is often overlooked, and it has helped to highlight the need for non-experts to have access to the methods being created within academia. Finally, whilst the types of tools developed can be used in a range of application areas, this research was concerned with applying FCA models to understand current and projected levels of chemotherapy provision within Wales, and with developing user-friendly tools that can help to assist key stakeholders in the planning and provision of resources.

In this chapter there is a reflection on the research that took place during the project, an evaluation of the aims and objectives which highlight the parts of the methodology which have attempted to address them. The contributions to knowledge are discussed highlighting the academic and policy implications. There is a section with some reflections of the project which use hindsight to focus on some areas where the project could have changed. The limitations of the project are explored and finally, there is a section detailing some future work which this project has highlighted.

8.2 Reflections on research aims and objectives

The aim of this PhD was to develop a free and open source tool which can be used by non-experts to model accessibility to healthcare (chemotherapy) services and provide meaningful insights into the provision of cancer services in Wales. This was achieved by providing a tool which was developed with the end user at the heart of the design. The tool is capable of multiple methods and it produces results that can be visualised very simply. It has been used to provide a current overview of the chemotherapy services within Wales, as well as providing a case study for alternative locations for additional resource.

The first objective of this project was to examine the uses of the 2SFCA method within healthcare and specifically within cancer related studies. This was achieved through the literature review which showed the different progressions within the 2SFCA methods and how these were used within different healthcare settings. Section 2.8 focused on different studies which were used specifically to investigate the management, treatment and planning of cancer services in different nations throughout the world.

Another objective was to investigate and evaluate the development of FOSS solutions that could be utilised to plan the provision of healthcare services. Section 4.2.1 investigated many of the available FOSS solutions available at the time of writing. This overview shows the main strengths and weaknesses of the different solutions with a particular focus on their routing and data handling capabilities. As well as this overview, the project built two tools capable of healthcare accessibility planning using two different FOSS solutions, and compared the strengths and weaknesses of the two platforms.

The need to collate and analyse a number of datasets which could be used to give a meaningful insight into the current provision of chemotherapy services within Wales was an objective which was met by compiling supply side data (Tenovus, FOI requests), demand side data (LSOA data, Tenovus user data) and network data. The data collected for this project gave an overview of the current chemotherapy provision within Wales where this data had not been publicly available.

The third objective was to illustrate the practical application of these tools using data on the current availability of cancer services in Wales through implementations based on FCA solutions. Using the tools developed it was possible to provide an up to date overview of chemotherapy accessibility in Wales, this was done using multiple methods from the 2SFCA family, Gaussian 2SFCA, E2SFCA (linear) and 2SFCA.

The fourth objective was to demonstrate the use of GIS-based accessibility models to examine the supply of mobile services in relation to potential demand for cancer services. This was achieved through the creation of multiple accessibility maps using Tenovus data to provide a detailed understanding of their current services, and also to suggest alternative locations for services such as chemotherapy mobile units, lymphoedema mobile units, choir sites, retail units and outreach services.

The aim of developing user-friendly tools to enable non-specialists to explore the implications of alternative models of provision of mobile services objective, and the test the usability of these tools with professionals concerned with helping to deliver cancer services objective were met by putting the user at the centre of the development process. Using a human computer interaction model, and utilising multiple methods with the Tenovus staff, it was possible to gather data on 3 iterations of the tool and refine the process, language and appearance of the tool so that it is usable by non-experts.

The final objective to draw on these findings to show how such tools can be used by the profession to plan and manage the provision of a wider range of healthcare services was met by completing a case study showcasing the capabilities of the tool with current chemotherapy data in Wales. The case study showed the possibility of adding additional or moving resources and re-computing the scores to understand the effects that these services are having in the different communities.

This project contributed to the collective knowledge in several ways. Many aspects of this project were approached from a different view point than they may normally be, and this has helped to provide some interesting contributions. By focusing on usability and not necessarily the ultimate algorithm for a specific data set, many of the findings should provide a good starting point for FOSS GIS projects and healthcare accessibility in practice.

There is great potential for this work to assist NHS Wales to reach one of their key actions highlighted in The Cancer Delivery Plan for Wales. Key action 40 (p. 11) which is “Where possible health boards should provide care locally and support patients who need assistance to travel or stay away from home.” The software solution developed during the course of this study has the capacity to be used to assist in measuring the current accessibility throughout Wales, but also to assist in the planning of additional services to maximise the equity of accessibility.

This tool has the ability to assist with advancing several of the policies laid out by Wales Cancer Network. It has the potential to assist with number 21 of the key actions set out in p.11, Cancer Delivery Plan for Wales 2016-2020 by helping policy makers understand the current level of accessibility within Wales; and also, as a tool to highlight the implications of additional services (see section 6.2). This also assists with several of the key points highlighted by charities such as The Wales Cancer Alliance priority policy calls which aim to address inequity in access throughout Wales.

An overview of the major FOSS GIS routing capabilities and open data allowed for future works to get a base understanding of the different programs and data, without needing to conduct the research on all the platforms. It was possible to pick some of the programs that are most closely aligned and then complete additional research on those specific platforms. There had previously been research in this area but it was more general and had not looked as closely at the routing capabilities of these platforms.

This project showed that at practical levels of use the difference between connecting a datapoint to a node or the nearest position on a line of a road network has minimal benefit at country scale, but can dramatically increase the complexity of the program; and therefore, the time that this takes to compute. For many instances of large-scale accessibility planning the difference is visually negligible and this knowledge can assist future research in managing time.

The work in this thesis highlighted the current state of chemotherapy accessibility in Wales which was previously unavailable at this scale. It used this data to share with key stakeholders what possible changes could affect large areas of Wales. This study had access to demand side data and was able to show the patterns of chemotherapy users to establish that the data is Gaussian in nature and provide a more precise catchment area for future studies. There are relatively few studies that have access to real demand data at this scale, and this has assisted in giving many future studies an ability to use a real metric as a catchment area, as opposed to an arbitrary distance or time.

8.3 Contributions to knowledge

This project has multiple key findings (section 7.3) and has made contributions to knowledge in several ways. It has helped to answer multiple policy actions and has added to the knowledge with several advancements within healthcare accessibility, usability studies and computing. The following contributions have been made.

A network comparison between OSM and OS Open Roads has been completed which shows the difference in performance between OSM and OS Open Roads compared to Google Maps sections 4.5.4, 6.4.2). The analysis used different length journeys within Wales and shows the differences between the datasets. The findings show the differences between ease of implementation and accuracy with OSM being closer to that of Google Maps and OS Open Roads being more straight forward to implement.

A FOSS GIS comparison was conducted which focussed on the different capabilities of the different options available (section 4.2.1). This comparison has added to the knowledge by having a source of information focussing on FOSS GIS and network analyst capabilities and will assist future researchers to understand the options available to them within this space.

A comparison of different routing options was completed, testing the differences in network performance with the different options of connecting to the nearest node or the nearest line (sections 4.3.2.4, 6.4.3). This work shows that connecting to the line is more computationally demanding but provides more realistic results although connecting to the nearest line is no guarantee that the node connects to the actual line that it should. The results showed that where possible connecting to the nearest line is optimal.

This project has continued the work of several previous papers by using accurate demand side data derived from Tenovus Cancer Care (Luo, 2014, Dai, 2011) and by conducting national scale cancer analysis (Shi et al., 2012). The use of this data provides a more accurate catchment size setting and allows for more confidence in the results.

User experience studies have been conducted with people involved in cancer service provision on FOSS GIS (section 5.5). There is little literature on FOSS GIS usability studies and also for academic user studies to be completed with cancer charities. These studies show how it is possible to keep the user at the centre of the development cycle whilst using tried and tested methods from usability literature.

This project compared the E2SFCA method using linear and Gaussian distance decay functions and compared these with the results of the original 2SFCA model. The results showed that the more advanced methods provided more realistic results but that even the 2SFCA method had the ability to highlight areas of poor accessibility.

These contributions will assist researchers to be more agile in the development of FOSS solutions as well as providing real world demand side data directly gathered from the source. By using the details within the FOSS GIS comparison table with the results of the network comparison and the user studies it should provide a solid foundation for future work. The use of the demand side data will enable cancer researchers in Wales to be more confident in their catchment sizes and provide further insight.

8.4 Limitations

There are limitations to the work completed during this thesis, and many lessons to be learned from completing a project of this type. There were technical challenges that were faced when using advanced methods and creating a tool for non-experts, and several technical limitations within this project. Many of these arose from the compromise between performance and complexity with usability. The focus of the project was not to make the fastest or most advanced tool but to make a usable tool that can be used in real world situations to help make decisions on service allocation.

The FCA methods used in this tool are not the most complete, there were several different methods that could have been optimised for this project, but these would have required increased complexity in terms of data preparation, computationally and operationally. Compromising on the FCA methods implemented in the tool allowed for advanced methods to be used, but whilst keeping a degree of usability which several of the alternative methods would not have allowed for (see table 2.1 and 2.2). Many of these proposed methods were designed for one type of dataset (Kim, Byon and Yeo, 2018) and to address one limitation (catchment size (Luo and Whippo, 2012), and different transport types (Langford, Higgs and Fry, 2016). This tool is designed to be used by multiple organisations across a range of data sets therefore optimisation of a method for a particular dataset is beyond the scope of the project.

This could be addressed by future users enabling more functionality within the tool and expanding upon the base that has been built with this program. By future users customising and adding additional functionality it could follow the same path as many open source software projects which get built upon by users willing to contribute to the project.

The way in which the results were visualised could have been more streamlined, and by having the tool in a GIS which uses a spatial database, or works alongside one would have assisted the user in only interacting with one GUI. Or it is possible to create a program from the ground up which would produce these visualisations. The advantages of QGIS for a company like Tenovus were huge and the security of a stand-alone spatial database mitigated the extra step in visualisation. However, this could be made a more streamlined process in the future if someone was able to work with QGIS and set up a process which is able to take the data and produce the map in the same way, there are also options in other FOSS GIS which could make this process more streamlined.

The scale of the analysis was a limitation in this project and more analysis could have taken place using different methods. Several more FCA methods could have been implemented to compare and contrast generated findings with those highlighted in this research. Trying to keep the user at the centre of the design and minimising complexity stopped this from being as expansive as it could have been, and more analysis would have assisted in understanding the effects of different methods on the data gathered.

The scale comparison could have included a larger range of data, and although the investigation highlighted that the finest data available provides the best results, it would have been of interest to show at what scale the data results in outcomes that offer little extra insight for the amount of increased effort required in data collation and analysis. Understanding the scale of the data required allows the tool to be used operationally with the data that they have available. Understanding that MSOA data is not as desirable as OA data could assist the users in data gathering and result in more detailed visualisations.

Due to the limits of Tenovus and the size of facility needed for their mobile units the case study was small. More consideration could have gone into alternative sites for the case study to provide more alternative locations for study and it may have been advantageous to identify new types of location which have not historically been used by Tenovus. A large amount of time was spent analysing data and ensuring the data produced insightful and actionable results, and this resulted in some of the empirical work having less detail than it could have.

This project was limited by the availability of complete data at suitable scales which could be used, supply side data was incredibly difficult to obtain. There is no known central place or organisation with data on where chemotherapy is provided within Wales, and the available data did not match between the different health boards. This made trying to understand the effects that Tenovus had throughout Wales very difficult, and with the FOI requests meant that output from this study is showing data that is not typically available.

The multiple FOI requests during this project caused considerable delay to delivering quality outputs. The requests were made to each health board with a variety of metrics required; this provided a good amount of choice when defining the supply side variable. The project mainly used the number of hours each unit is open per week multiplied by the number of chairs or beds available for treatment. This helped to stop the results being skewed by a 24-hour ward or a large ward offering services 2 hours a week (see section 4.3.2).

Before the FOI requests were made there were multiple contacts with many different industry bodies with, and without the aid of Tenovus to get this information. Many people offered to assist, but there were never any results from these conversations. There was a large amount of time spent investigating each health board to understand their offerings, but from the available resources online it was almost impossible to decipher the locations at which chemotherapy was offered.

As well as the difficulties in obtaining accurate supply side data, the open network data varied in quality dramatically depending on where it was sourced. Network data is very large and complex in nature and the quality of it can have a large effect upon the results by skewing the sizes of the catchments created. Keeping to the open data meant that the data available was limited, the open street map dataset was very detailed and the OS Open Roads dataset was more limited in its feature set. Ideally it would have been good to connect to an API and use something like Google maps to provide the routing functions, but this could have become quite expensive very quickly. It would have removed a lot of the complexity from the super user which is something that any potential user could investigate moving forward. The full Ordnance Survey paid for data set is also of good quality and could be considered by an organisation which had access to it.

Overall, it was very difficult to get data at a scale suitable for this project and this difficulty could make utilisation of the tool by organisations less frequent. It was noted that the FOI requests worked well, and would have been completed much earlier in if the project was to be undertaken again. It would also be important to gain additional data in the neighbouring countries such as England to provide a much clearer picture of the border regions. Finally, as chemotherapy varies hugely, it would have been good to gain a richer understanding of the different types of treatments and the locations at which these were provided, so that an analysis of different types of cancer treatments could have been conducted.

This could be mitigated in future by setting up in house processes to capture the relevant data and store it in the correct format and by building relationships with the health boards and requesting the same data from each one. The relationships that Tenovus has with the health boards could assist in providing a standard of reporting to assist with chemotherapy, and potentially other areas of interest moving forward.

There were several organisational limitations which need to be addressed by Tenovus before the tool can be fully implemented. The need for a super user needs to be discussed internally and the competent user needs to spend some time either reading the documentation or having a training session where the installations can take place and the database can be set up. Decisions will have to be made within Tenovus on how they implement this; whether it is on one machine, or whether it is rolled out to desktop throughout the organisation.

Tenovus also needs to decide how and when the tool is used and there needs to be a conversation between the key stakeholders to ensure that it is discussed during planning meetings. The potential this tool has to assist them and others in planning should be highlighted and also discussed with external organisations who can make use of this (cancer screening charities, health boards). There is a learning curve involved with this tool, and some staff may be reluctant to take on additional tasks, or ones where they need to spend some time learning how unfamiliar systems work. There is often resistance to change, and some staff may be hesitant to change how they have been working previously. This learning curve has been mitigated as far as possible without losing vital functionality, but it could still affect potential usage.

One of the more notable results of the user studies was that the participant did not necessarily understand the potential of these methods and how they could be used within their role. This was mitigated within Tenovus by using examples close to the work of the participant and having assistance from the management team; in other organisations it may be more difficult to introduce these measures from the top down without first helping them understand the potential. Although this tool provides advanced accessibility measures it should only be one part of a process used to understand accessibility. Other factors remain such as the distance different nurses are able to travel to work on the unit, and the lack of a facility to host the mobile unit. The health boards also have a considerable say in this process, and they may limit the movement of the units based on their boundaries. Tenovus also provides many other services for which this tool may not be the most suitable accessibility measure (charity shops), as the method has been optimised for healthcare accessibility, not retail.

8.5 Future Directions

A number of limitations were identified throughout the research leading to potential for further research development. These limitations are;

Application to a wider range of scenarios

The tool could be used in many other ways, it would be interesting to apply these methods in other scenarios such as financial support for cancer services, or psychological counselling for the recently bereaved. These services are offered by Tenovus and other organisations, and optimising the tool for these uses would help create a more useful tool which could be of use to more people.

Research on Utility

It would be of interest to understand how Tenovus and other organisations actually use tools such as this. With this information it would be possible to create a better design brief for future projects. Understanding how these tools are used and what the outputs are used for in reality would be valuable to academia and policy makers when deciding how best to move this research from an academic proposition to one which is used more regularly in practice.

Advancements in cancer applications

There are constant improvements in how cancer is treated and this is done with many different treatment types. It would be of interest to evaluate these different methods and establish if they could be used to plan how and where these treatments could be positioned from an accessibility viewpoint. By optimising the tool for different treatments, it could be possible to provide real input to the policy makers and assist in better all-round cancer treatment. This could build on previous research that has examined potential relationships between access to cancer services and healthcare outcomes, or the stage of diagnosis of cancer so as to examine the potential impacts of inequalities in service provision.

Developments within FCA

FCA methods are increasingly being extended both from methodological and thematic standpoints, and there is room to evaluate the different methods and provide either a framework or additional methods to the tool which can be used by treatment providers and policy makers alike. Having the knowledge of which method should be used for a particular type of analysis would assist in the efficacy of the tool and could be useful outside of cancer provision as well.

Methodology

Throughout this project there has been a strong emphasis on providing complex methods to non-experts who are actually working in these fields. A number of the different accessibility methods proposed are technically more rounded, or deal with a particular problem better, but many of these improvements are making the methodology more complicated and relying on complex specialist computer programming; and many organisations simply cannot or will not pay for this type of analysis to take place. Some analysis should be conducted using standard data to understand the level of benefit these methodologies offer, and look at ways to simplify the process to increase the use of the methods which can assist in real world change.

Human Computer Interaction

There is minimal work in the human computer interaction focusing on FOSS or GIS. This is something that could assist in the growth of FOSS utilisation in non-expert settings and would enable potential users to have a better experience. It would be interesting to find out how much resource projects allow for user studies, and if they use any particular models, and then finally the production of a specific model for FOSS systems focusing on the unique challenges of FOSS development and laying out a foundation for them to follow.

Data

Much work has been done investigating the supply side data for healthcare accessibility, but there is very limited work investigating the demand side travel time; in many studies arbitrary numbers are used to propose travel time or distance. The data for this can be very sensitive and difficult to obtain, but having knowledge of how healthcare users travel to a service and the times taken would allow for more accurate findings from studies.

This study gathered the chemotherapy and lymphoedema provisions within Wales to provide an accurate overview of cancer provisions, but as cancer is complex there are many patients who have travelled elsewhere within the UK to get treatment. It would be of great benefit to understand how far individuals are travelling to access a wider range of cancer services offered by other providers, and obtain data on which patients get treatment at which facilities. Strict ethical guidelines often prohibit the use of such datasets, but this study has shown how they could potentially be used to inform the use of accessibility models. This would help provide a more complete picture of the cancer provisions within the UK, and highlight the outlier patients who are required to travel many miles for treatment.

Finally, this project has been interested in demonstrating at what scales different elements of the methodology have an impact. It would be interesting to see this developed further to establish at what point the quality of data has diminishing returns, and whether this data can provide misleading results if it is not implemented correctly. By investigating the effects, the data has upon the methodology it is possible to get a more detailed understanding of the results and assist in collecting correct data at the appropriate scale in relation to project objectives.

8.6 Policy implications

The cancer delivery plan for Wales (Cancer Delivery Plan for Wales 2016-2020) highlights several areas. In service delivery the plan suggests that “Services should meet the different needs of the population” and “The whole issue of local access to services is one which must form a delicate balance between accessibility and access to high quality, sustainable services” (p.10, Cancer Delivery Plan for Wales 2016-2020).

Key action 21 (page 12) is “The regional cancer centres, Wales Cancer Network and WHSSC to work together to ensure equity of access and delivery of service quality”. It is clear that access to services is a priority within Wales, and having access to a free and open source tool which can be used by non-experts to assist in the delivery of this could be of great benefit to achieving this action. It is possible for key stakeholders to work individually or across the Welsh healthcare spectrum to better understand the current state of chemotherapy provision within Wales, but to also address the issues with cancer screening and after care services.

Key action 31 (page 12) is “Health boards to continue to monitor performance against national waiting times targets and focus on improving performance”. This key action can be assisted by having further understanding of the current provisions, and this tool could be utilised to plan extra capacity in the areas of greatest need. It can be a useful tool to add further knowledge which can be assessed by the key stakeholders.

Key action 32 (page 13) is “Health boards to ensure cross border flows do not disadvantage patients in access to treatment and care.” Understanding the level of access within Wales and with the border provisions being shown, would allow the different health boards to gain a better knowledge of when and where people are crossing the border for treatment. This tool could provide assistance in the planning of provisions to mitigate any unnecessary cross border treatments that may occur.

Key action 40 (page 15) is “Where possible health boards should provide care locally and support patients who need assistance to travel or stay away from home.” This action is directly linked to this tool, and could be used by the different stakeholders to provide a current overview of the provisions as well as to assist in providing more cancer services closer to home.

This tool could be used for several of the Cancer Delivery Plan for Wales’ key actions, and is versatile enough to be of use in other areas of healthcare provision. It has the potential to affect and enhance the way in which NHS Wales plan and monitor their services.

Aim number 1 of Tenovus is “To provide support and treatment to people with cancer and their loved ones, closer to home in unique ways”[13] and this tool is designed to assist them in being more effective with this aim. The ability to understand the effect that their service has on the overall chemotherapy treatments within Wales enables them to be more targeted in areas where they have maximum impact. It is important for Tenovus to ensure that they deliver the best value from their donations, and being able to show supporters how they are affecting the chemotherapy delivery within Wales is of great help to them.

[13] https://www.tenovuscancercare.org.uk/about/about-us/

Tenovus have presented a manifesto “More than medicine” as the Welsh Assembly is being formed in 2021[14] and this project has the potential to assist in several key areas. One of the top priorities is to “Take treatment closer to home” where clinically possible using their mobile units. This tool is ideally positioned to assist Tenovus and the different stakeholders to understand the current state of cancer provisions within Wales but to also assist in the planning of additional provision. It has the potential to add some accuracy to the process which currently is not very scientific.

The manifesto also outlines the need for better data across Wales as a whole to allow health boards to be accountable (p. 13[15]). On many occasions during the course of this project data gathering was difficult and this was highlighted within Tenovus at many meetings. It is hoped that this project has had some influence in highlighting the difficulty in data gathering, and why it is important to have all the health boards reporting consistently.

Tenovus (p.16[16]) have highlighted that lung cancer could be screened for, and suggested targeted groups. The ability to use the WIMD data to precisely target different age and sex groups for screening locations would add huge value in the planning and execution of lung and other screening strategies.

This study has highlighted some gaps within the FCA research at present by focusing on usability and real-world uptake. It has highlighted that increasing complexity and methods are important, but it is just as important to make these methods available to those in industry who can affect change. This project has highlighted the need to investigate the ways in which these methods are being utilised, and has been able to show that patients receiving treatment in Wales travel within around 35km. The data is Gaussian in nature, this can be accessed by future studies to provide a more solid benchmark for the catchment areas being proposed.

---

[14] https://www.tenovuscancercare.org.uk/wp-content/uploads/2020/06/tenovus-cancer-care-more-than-medicine.pdf

[15] https://www.tenovuscancercare.org.uk/wp-content/uploads/2020/06/tenovus-cancer-care-more-than-medicine.pdf

[16] https://www.tenovuscancercare.org.uk/wp-content/uploads/2020/06/tenovus-cancer-care-more-than-medicine.pdf

8.7 Project reflections

Within the project and with hindsight there may have been some missed opportunities to try different solutions or take different routes. These decisions are difficult to understand in real time, but reflection allows for some questions to be asked. gvSIG or GRASS may have provided a better solution as they allow for an internal spatial database, and they have a network analyst function within them. The GRASS geodatabase could also have been accessed from QGIS and most of the functionality would have been kept within one platform. Whilst acknowledging the potential of such an approach, there would still be the need for the tool to be opened for information to be input, and it would not make the program quicker or any more intuitive. The fact that both gvSIG and GRASS are updated regularly would mean that a super user would have to be able to ensure the updates were backwards compatible, and potentially change parts of the program to keep it current.

Further network research with APIs would have been a useful addition to the project. The fact that it costs money to access these types of service meant they were discounted, but things like the Open Source Routing Machine could have been investigated more, and potentially having a server which could be used by those accessing the tool. Further investigation into this type of route provision would have been beneficial.

Additional user studies and different methods would have been interesting as there is very little user experience research within FOSS GIS; this could have been impactful and provided some meaningful results for the open source community. This is something that could have added to the project, but the results have assisted in showing how the non-expert user operates in these types of systems, and there is a reason why most people that use them have a considerable amount of training.

There was room to incorporate many additional types of E2SFCA models, highlighted in chapter 2, such as the 3-step model or those that adopt a multi-modal FCA approach. This would have made the tool more advanced, and increased its effectiveness at different scales and with different types of data. Increasing the functionality would have been a good way to give the user more and would have made the tool more versatile. The problem with this is that it would also increase the knowledge needed by the user, and make it more difficult to explain to them how it works. Many of the advancements within the FCA methods family have required additional special or aspatial data and additional inputs for them to work, which is fine in an academic or professional setting but does not translate well to non-expert users. The inclusion of the Gaussian method was only done as the unique demand side dataset highlighted that it was Gaussian in nature.

References

Adams, S. A., Choi, S. K., Khang, L., A. Campbell, D., Friedman, D. B., Eberth, J. M., Glasgow, R. E., Tucker-Seeley, R., Xirasagar, S., Yip, M. P., Young, V. M., & Hébert, J. R. (2015). Decreased Cancer Mortality-to-Incidence Ratios with Increased Accessibility of Federally Qualified Health Centers. *Journal of Community Health*, *40*(4), 633–641. https://doi.org/10.1007/s10900-014-9978-8.

Apparicio, P., Abdelmajid, M., Riva, M., & Shearmur, R. (2008). Comparing alternative approaches to measuring the geographical accessibility of urban health services: Distance types and aggregation-error issues. *International Journal of Health Geographics*, *7*(1), 7. https://doi.org/10.1186/1476-072X-7-7.

Jasiewicz, J., & Metz, M. (2011). A new GRASS GIS toolkit for Hortonian analysis of drainage networks. *Computers and Geosciences*, 1–12. https://doi.org/10.1016/j.cageo.2011.03.003.

Bauer, J., & Groneberg, D. A. (2016). Measuring spatial accessibility of health care providers-introduction of a variable distance decay function within the floating catchment area (FCA) method. *PLoS ONE*, *11*(7), 1–17. https://doi.org/10.1371/journal.pone.0159148.

Bauer, J., Groneberg, D. A., Maier, W., Manek, R., Louwen, F., & Brüggmann, D. (2017). Accessibility of general and specialized obstetric care providers in Germany and England: an analysis of location and neonatal outcome. *International Journal of Health Geographics*, *16*(1), 44. https://doi.org/10.1186/s12942-017-0116-6.

Bauer, J., Müller, R., Brüggmann, D., & Groneberg, D. A. (2018). Spatial Accessibility of Primary Care in England: A Cross-Sectional Study Using a Floating Catchment Area Method. *Health Services Research*, *53*(3), 1957–1978. https://doi.org/10.1111/1475-6773.12731.

Belasco, E. J., Gong, G., Pence, B., Wilkes, E., & Wilkes, E. (2014). The Impact of Rural Health Care Accessibility on Cancer-Related Behaviors and Outcomes. *Applied Health Economics and Health Policy*. https://doi.org/10.1007/s40258-014-0099-4.

Benson, C., Muller-Prove, M., & Mzourek, J. (2004). *Professional usability in open source projects*. 1083. https://doi.org/10.1145/985921.985991.

Bhaumik, C., Agrawal, A. K., Adak, S., Ghose, A., & Das, D. (2012). *Sensor Observation Service based Medical Instrument Integration Data Capture, Analysis and Rendering*.

Brandão, A., Pereira, E., Esteves, M., Portela, F., Santos, M. F., Abelha, A., & Machado, J. (2016). A benchmarking analysis of open-source business intelligence tools in healthcare environments. *Information (Switzerland)*, *7*(4), 1–16. https://doi.org/10.3390/info7040057.

Cabrera Barona, P., & Blaschke, T. (2015). Healthcare Accessibility and Socio-economic Deprivation: A Case Study in Quito, Ecuador. *GI_Forum - Journal for Geographic Information Science*, *1*, 484–492. https://doi.org/10.1553/giscience2015s484.

Cai, E., Liu, Y., Jing, Y., Zhang, L., Li, J., & Yin, C. (2017). Assessing spatial accessibility of public and private residential aged care facilities: A case study in Wuhan, Central China. *ISPRS International Journal of Geo-Information*, *6*(10). https://doi.org/10.3390/ijgi6100304.

Cancer Delivery Plan for Wales 2016-2020. (2016).

Cao, Y., Stewart, K., & Kalil, R. (2016). Geographic patterns of end-stage renal disease and kidney transplants in the Midwestern United States. *Applied Geography*, *71*, 133–143. https://doi.org/10.1016/j.apgeog.2016.05.001.

Carillo, K., & Okoli, C. (2008). *THE OPEN SOURCE MOVEMENT: A REVOLUTION IN SOFTWARE DEVELOPMENT*.

Chen, X. (2019). *The Professional Geographer Enhancing the Two-Step Floating Catchment Area Model for Community Food Access Mapping: Case of the Supplemental Nutrition Assistance Program*. https://doi.org/10.1080/00330124.2019.1578978.

Chen, Z., & Yeh, A. G. O. (2019). Accessibility Inequality and Income Disparity in Urban China: A Case Study of Guangzhou. *Annals of the American Association of Geographers*, *109*(1), 121–141. https://doi.org/10.1080/24694452.2018.1470923.

Choosumrong, S., Raghavan, V., & Bozon, N. (2012). Multi-Criteria Emergency Route Planning Based on Analytical Hierarchy Process and pgRouting. *Geoinformatics*, *23*(4), 159–167. https://doi.org/10.6010/geoinformatics.23.159.

Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., & Prior, F. (2013). The cancer imaging archive (TCIA): Maintaining and operating a public information repository. *Journal of Digital Imaging*, *26*(6), 1045–1057. https://doi.org/10.1007/s10278-013-9622-7.

Cristina Rocha-Brischiliari, S. I., Andrade, L., Kenji Nihei, O., Brischiliari, A., dos Santos Hortelan, M., Dalva de Barros Carvalho, M., & Marisa Pelloso, S. (2018). *Spatial distribution of breast cancer mortality: Socioeconomic disparities and access to treatment in the state of Parana, Brazil*. https://doi.org/10.1371/journal.pone.0205253.

Cromley, E. K., & McLafferty, S. L. (2002) *GIS and Public Health* (Guilford Publications, New York).

Dai, D. (2011). Racial/ethnic and socioeconomic disparities in urban green space accessibility: Where to intervene? *Landscape and Urban Planning*, *102*(4), 234–244. https://doi.org/10.1016/j.landurbplan.2011.05.002.

Dai, D., & Wang, F. (2011). Geographic disparities in accessibility to food stores in southwest Mississippi. *Environment and Planning B: Planning and Design*, *38*(4), 659–677. https://doi.org/10.1068/b36149.

Dai, T. qi, Liu, Z. bing, Liao, C., & Cai, H. yu. (2018). Incorporating job diversity preference into measuring job accessibility. *Cities*, *78* (October 2017), 108–115. https://doi.org/10.1016/j.cities.2018.02.003.

Dewulf, B., Neutens, T., De Weerdt, Y., & Van De Weghe, N. (2013). Accessibility to primary health care in Belgium: An evaluation of policies awarding financial assistance in shortage areas. *BMC Family Practice*, *14*(1), 1. https://doi.org/10.1186/1471-2296-14-122.

Díaz, L., & Anguix, A. (2008). gvSIG: A GIS desktop solution for an open SDI. In *Journal of Geography and Regional Planning* (Vol. 1, Issue 3). http://www.academicjournals.org/JGRP.

Donohoe, J., Marshall, V., Tan, X., Camacho, F. T., Anderson, R., & Balkrishnan, R. (2015). Predicting late-stage breast cancer diagnosis and receipt of adjuvant therapy applying current spatial access to care methods in appalachia. *Medical Care*, *53*(11), 980–988. https://doi.org/10.1097/MLR.0000000000000432.

Donohoe, J., Marshall, V., Tan, X., Camacho, F. T., Anderson, R., & Balkrishnan, R. (2016). Evaluating and comparing methods for measuring spatial access to mammography centers in Appalachia. *Health Services and Outcomes Research Methodology*, *16*(1–2), 22–40. https://doi.org/10.1007/s10742-016-0143-y.

Dorner, W., Scheffer, J., & Zink, R. (2011). *GvSIG in the academic education of heterogeneous target groups-experiences in lectures, exercises and eLearning*.

Eduardo Oliveira Lizardo, L., Augusto Davis Jr, C., & Eduardo Oliveira Lizardo, L. (2017). *A PostGIS extension to support advanced spatial data types and integrity constraints*. *10*. https://doi.org/10.1145/3139958.3140020.

Elhag, H. M. A., & Abushama, H. M. (2011). *Migration to FOSS : Readiness and Challenges*.

Elisabeth Hove, S., & Anda, B. (2005). *Experiences from Conducting Semi-Structured Interviews in Empirical Software Engineering Research*.

Fitzgerald, B. (2006). *The Transformation of Open Source Software SCALARE: Scaling Software View project Software Development with Alternative Workforces (SODAW) View project THE TRANSFORMATION OF OPEN SOURCE SOFTWARE 1*. *30*(3), 587–598. https://doi.org/10.2307/25148740.

Fransen, K., Neutens, T., De Maeyer, P., & Deruyter, G. (2015). A commuter-based two-step floating catchment area method for measuring spatial accessibility of daycare centers. *Health and Place*, *32*, 65–73. https://doi.org/10.1016/j.healthplace.2015.01.002.

Ganesh, G. R. D., Jaidurgamohan, K., Srinu, V., Kancharla, C. R., & Suresh, S. V. S. (2016). Design of a low cost smart chair for telemedicine and IoT based health monitoring: An open source technology to facilitate better healthcare. *11th International Conference on Industrial and Information Systems, ICIIS 2016 - Conference Proceedings*, *2018-Janua*, 89–94. https://doi.org/10.1109/ICIINFS.2016.8262913.

Gao, F., Kihal, W., Le Meur, N., Souris, M., & Deguen, S. (2016). Assessment of the spatial accessibility to health professionals at French census block level. *International Journal for Equity in Health*, *15*(1), 1–14. https://doi.org/10.1186/s12939-016-0411-z.

Garrett, J. J. (2011). *The Elements of User Experience: User-Centered Design for the Web and Beyond, Second Edition Notice of Rights Notice of Liability*. www.newriders.com.

Gilliland, J. A., Shah, T. I., Clark, A., Sibbald, S., & Seabrook, J. A. (2019). A geospatial approach to understanding inequalities in accessibility to primary care among vulnerable populations. *PLoS ONE*, *14*(1). https://doi.org/10.1371/journal.pone.0210113.

Gorey, K. M., Luginaah, I. N., Bartfay, E., Fung, K. Y., Holowaty, E. J., Wright, F. C., Hamm, C., & Kanjeekal, S. M. (2011). Effects of socioeconomic status on colon cancer treatment accessibility and survival in Toronto, Ontario, and San Francisco, California, 1996-2006. *American Journal of Public Health*, *101*(1), 112–119. https://doi.org/10.2105/AJPH.2009.173112.

Gu, W., Wang, X., & McGregor, S. E. (2010). Optimization of preventive health care facility locations. *International Journal of Health Geographics*, *9*(1), 17. https://doi.org/10.1186/1476-072X-9-17.

Gu, X., Tao, S., & Dai, B. (2017). Spatial accessibility of country parks in Shanghai, China. *Urban Forestry and Urban Greening*, *27* (December 2016), 373–382. https://doi.org/10.1016/j.ufug.2017.08.006.

Guo, Y., Chan, C. H., & Yip, P. S. F. (2017). Spatial variation in accessibility of libraries in Hong Kong. *Library and Information Science Research*, *39*(4), 319–329.

https://doi.org/10.1016/j.lisr.2017.11.007.

Higgs, G., Langford, M., Jarvis, P., Page, N., Richards, J., & Fry, R. (2019). Using Geographic Information Systems to investigate variations in accessibility to ‘extended hours’ primary healthcare provision. *Health and Social Care in the Community*, *27*(4), 1074–1084. https://doi.org/10.1111/hsc.12724.

Jamtsho, S., Corner, R., & Dewan, A. (2015). Spatio-Temporal Analysis of Spatial Accessibility to Primary Health Care in Bhutan. *ISPRS International Journal of Geo-Information*, *4*(3), 1584–1604. https://doi.org/10.3390/ijgi4031584.

Jin, M., Liu, L., Tong, D., Gong, Y., & Liu, Y. (2019). Evaluating the spatial accessibility and distribution balance of multi-level medical service facilities. *International Journal of Environmental Research and Public Health*, *16*(7). https://doi.org/10.3390/ijerph16071150.

Jones, M., Shah, T., Hayes, A., Uswak, G., & Bell, S. (2016). Dental service disparities in Canada: A Saskatoon, Saskatchewan case study. *Canadian Geographer*, *60*(3), 346–355. https://doi.org/10.1111/cag.12286.

Joseph, A. E., & Bantock, P. R. (1982). Measuring potential physical accessibility to general practitioners in rural areas: A method and case study. *Social Science and Medicine*, *16*(1), 85–90. https://doi.org/10.1016/0277-9536(82)90428-2.

Jung, M. (2013). *LecoS-A QGIS plugin for automated landscape ecology analysis*. https://doi.org/10.7287/peerj.preprints.116v2.

Kanuganti, S., Sarkar, A. K., & Singh, A. P. (2016). Evaluation of access to health care in rural areas using enhanced two-step floating catchment area (E2SFCA) method. *Journal of Transport Geography*, *56*, 45–52. https://doi.org/10.1016/j.jtrangeo.2016.08.011.

KC, K., Corcoran, J., & Chhetri, P. (2018). Measuring the spatial accessibility to fire stations using enhanced floating catchment method. *Socio-Economic Planning Sciences*. https://doi.org/10.1016/J.SEPS.2018.11.010.

Kelly, C., Hulme, C., Farragher, T., & Clarke, G. (2016). Are differences in travel time or distance to healthcare for adults in global north countries associated with an impact on health outcomes? A systematic review. In *BMJ Open* (Vol. 6, Issue 11, p. e013059). BMJ Publishing Group. https://doi.org/10.1136/bmjopen-2016-013059.

Khakh, A. K., Fast, V., & Shahid, R. (2019). Spatial accessibility to primary healthcare services by multimodal means of travel: Synthesis and case study in the city of calgary. *International Journal of Environmental Research and Public Health*, *16*(2). https://doi.org/10.3390/ijerph16020170.

Khan-Gates, J. A., Ersek, J. L., Eberth, J. M., Adams, S. A., & Pruitt, S. L. (2015). Geographic Access to Mammography and Its Relationship to Breast Cancer Screening and Stage at Diagnosis: A Systematic Review. *Women's Health Issues*, *25*(5), 482–493. https://doi.org/10.1016/j.whi.2015.05.010.

Khan, A. A. (1992). An Integrated Approach to Measuring Potential Spatial Access to Health Care Services. In *Plann. Sci* (Vol. 26, Issue 4).

Khan, S., & Mutahar Aaqib, S. (2017). *Empirical Evaluation of ArcGIS with Contemporary Open Source Solutions-A Study*. www.fsf.org.

Kim, Y., Byon, Y.-J., & Yeo, H. (2018). Enhancing healthcare accessibility measurements using GIS: A case study in Seoul, Korea. *PLOS ONE*, *13*(2), e0193013. https://doi.org/10.1371/journal.pone.0193013.

Kocatepe, A., Baran Ulak, M., Ozguven, E., Horner, M. W., & Arghandeh, R. (2017). *Socioeconomic characteristics and crash injury exposure: A case study in Florida using two-step floating catchment area method*. https://doi.org/10.1016/j.apgeog.2017.08.005.

Kucuk, A., Hamdi, S. M., Aydin, B., Schuh, M., Hamdi, M., Schuh, M. A., & Angryk, R. A. (2016). *Pg-Trajectory: A PostgreSQL/PostGIS Based Data Model for Spatiotemporal Trajectories*. https://doi.org/10.1109/BDCloud-SocialCom-SustainCom.2016.23.

Langford, M., Fry, R., & Higgs, G. (2012). Measuring transit system accessibility using a modified two-step floating catchment technique. *International Journal of Geographical Information Science*, *26*(2), 193–214. https://doi.org/10.1080/13658816.2011.574140.

Langford, Mitchel, Higgs, G., & Fry, R. (2016). Multi-modal two-step floating catchment area analysis of primary health care accessibility. *Health and Place*, *38*, 70–81. https://doi.org/10.1016/j.healthplace.2015.11.007.

Langford, Mitchel, Higgs, G., & Radcliffe, J. (2018). The application of network-based GIS tools to investigate spatial variations in the provision of sporting facilities. *Annals of Leisure Research*, *21*(2), 178–198. https://doi.org/10.1080/11745398.2016.1272059.

Leal, W., Belay, F., Jokasha, S., Menas, K., Pantaleo, W., Kumbirai, M., & Editors, M. (2017). *Climate Change Management Climate Change Adaptation in Africa Fostering Resilience and Capacity to Adapt*. Retrieved February 12, 2020, from http://www.springer.com/series/8740.

Ledo, D., Houben, S., Vermeulen, J., Marquardt, N., Oehlberg, L., & Greenberg, S. (2018). Evaluation strategies for HCI Toolkit research. In *Conference on Human Factors in Computing Systems - Proceedings* (Vols. 2018-April). https://doi.org/10.1145/3173574.3173610.

Leech, B. L. (2002). Asking questions: Techniques for semistructured interviews. *PS - Political Science and Politics*, *35*(4), 665–668. https://doi.org/10.1017/S1049096502001129.

Li, L., Du, Q., Ren, F., & Ma, X. (2019). Assessing spatial accessibility to hierarchical urban parks by multi-types of travel distance in Shenzhen, China. *International Journal of Environmental Research and Public Health*, *16*(6). https://doi.org/10.3390/ijerph16061038.

Lin, B. C., Chen, C. W., Chen, C. C., Kuo, C. L., Fan, I. chun, Ho, C. K., Liu, I. C., & Chan, T. C. (2016). Spatial decision on allocating automated external defibrillators (AED) in communities by multi-criterion two-step floating catchment area (MC2SFCA). *International Journal of Health Geographics*, *15*(1), 1–14. https://doi.org/10.1186/s12942-016-0046-8.

Luo, J. (2014). Integrating the huff model and floating catchment area methods to analyze spatial access to healthcare services. *Transactions in GIS*, *18*(3), 436–448. https://doi.org/10.1111/tgis.12096.

Luo, J. (2016). Analyzing Potential Spatial Access to Primary Care Services with an Enhanced Floating Catchment Area Method. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *51*(1), 12–24. https://doi.org/10.3138/cart.51.1.3230.

Luo, W. (2004). Using a GIS-based floating catchment method to assess areas with shortage of physicians. *Health and Place*, *10*(1), 1–11. https://doi.org/10.1016/S1353-8292(02)00067-9.

Luo, W., & Qi, Y. (2009). An enhanced two-step floating catchment area (E2SFCA) method for measuring spatial accessibility to primary care physicians. *Health and Place*, *15*(4), 1100–1107. https://doi.org/10.1016/j.healthplace.2009.06.002.

Luo, W., & Wang, F. (2003). Measures of spatial accessibility to health care in a GIS environment: Synthesis and a case study in the Chicago region. *Environment and Planning B: Planning and Design*, *30*(6), 865–884. https://doi.org/10.1068/b29120.

Luo, W., & Whippo, T. (2012). Variable catchment sizes for the two-step floating catchment area (2SFCA) method. *Health and Place*, *18*(4), 789–795. https://doi.org/10.1016/j.healthplace.2012.04.002.

Lyimo, F. S., & Beran, T. N. (2012). Demographic, knowledge, attitudinal, and accessibility factors associated with uptake of cervical cancer screening among women in a rural district of Tanzania: Three public policy implications. *BMC Public Health*, *12*(1), 1–8. https://doi.org/10.1186/1471-2458-12-22.

Ma, L., Luo, N., Wan, T., Hu, C., & Peng, M. (2018). An improved healthcare accessibility measure considering the temporal dimension and population demand of different ages. *International Journal of Environmental Research and Public Health*, *15*(11). https://doi.org/10.3390/ijerph15112421.

Mao, L., & Nekorchuk, D. (2013). Measuring spatial accessibility to healthcare for populations with multiple transportation modes. *Health and Place*, *24*, 115–122. https://doi.org/10.1016/j.healthplace.2013.08.008.

McGrail, M. R., & Humphreys, J. S. (2009a). Measuring spatial accessibility to primary care in rural areas: Improving the effectiveness of the two-step floating catchment area method. *Applied Geography*, *29*(4), 533–541. https://doi.org/10.1016/j.apgeog.2008.12.003.

McGrail, M. R., & Humphreys, J. S. (2009b). The index of rural access: An innovative integrated approach for measuring primary care access. *BMC Health Services Research*, *9*, 1–12. https://doi.org/10.1186/1472-6963-9-124.

McGrail, M. R., & Humphreys, J. S. (2014). Measuring spatial accessibility to primary health care services: Utilising dynamic catchment sizes. *Applied Geography*, *54*, 182–188. https://doi.org/10.1016/j.apgeog.2014.08.005.

McLafferty, S., & Wang, F. (2009). Rural reversal? *Cancer*, *115*(12), 2755–2764. https://doi.org/10.1002/cncr.24306.

McLafferty, S., Wang, F., Luo, L., & Butler, J. (2011). Rural - urban inequalities in late-stage breast cancer: Spatial and social dimensions of risk and access. *Environment and Planning B: Planning and Design*, *38*(4), 726–740. https://doi.org/10.1068/b36145.

Mooney, P., Corcoran, P., & Winstanley, A. C. (2010). *Towards Quality Metrics for OpenStreetMap*.

Naranjo-Gómez, J. M., Castanho, R. A., Cabezas-Fernández, J., & Loures, L. C. (2019). Esta obra se distribuye con la licencia Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional Assessing High Speed Railway Service Areas in Spain Mainland from a Multi-method GIS Approach. *Revista de Estudios Andaluces*, *37*, 2340–2776. https://doi.org/10.12795/rea.2019.i37.09.

Nassar, V. (2012). *Common criteria for usability review*. https://doi.org/10.3233/WOR-2012-0282-1053.

Nasseh, K., Eisenberg, Y., & Vujicic, M. (2017). Geographic access to dental care varies in Missouri and Wisconsin. *Journal of Public Health Dentistry*, *77*(3), 197–206. https://doi.org/10.1111/jphd.12197.

Naylor, K. B., Tootoo, J., Yakusheva, O., Shipman, S. A., Bynum, J. P. W., & Davis, M. A. (2019). Geographic variation in spatial accessibility of U.S. healthcare providers. *PLOS ONE*, *14*(4), e0215016. https://doi.org/10.1371/journal.pone.0215016.

Ngamini Ngui, A., & Vanasse, A. (2012). Assessing spatial accessibility to mental health facilities in an urban environment. *Spatial and Spatio-Temporal Epidemiology*, *3*(3), 195–203. https://doi.org/10.1016/j.sste.2011.11.001.

Ngui, A. N., & Apparicio, P. (2011). Optimizing the two-step floating catchment area method for measuring spatial accessibility to medical clinics in Montreal. *BMC Health Services Research*, *11*(1), 166. https://doi.org/10.1186/1472-6963-11-166.

Ni, J., Wang, J., Rui, Y., Qian, T., & Wang, J. (2015). An enhanced variable two-step floating catchment area method for measuring spatial accessibility to residential care facilities in Nanjing. *International Journal of Environmental Research and Public Health*, *12*(11), 14490–14504. https://doi.org/10.3390/ijerph121114490.

Nielsen, A., Bolding, K., Hu, F., & Trolle, D. (2017). An open source QGIS-based workflow for model application and experimentation with aquatic ecosystems. *Environmental Modelling and Software*, *95*, 358–364. https://doi.org/10.1016/j.envsoft.2017.06.032.

Nielsen, J, & Landauer, J. (1993). A mathematical model of finding the usability problem. Proceedings of the CHI 93 proceedings of the Interact conference on human factors in computing systems. *Proceedings of ACM INTERCHI'93 Conference*, 206–213. https://doi.org/10.1145/169059.169166.

Nielsen, Jakob. (1997). The Use and Misuse of Focus Groups. *IEEE Software*, *14*(1), 94–95. https://doi.org/10.1109/52.566434.

Norman, D. A. (1988). *The design of everyday things*. Doubleday.

Norman, D. A., & Draper, S. W. (1986). *User Centered System Design: new perspectives on Human-Computer Interaction.*

Pan, J., Liu, H., Wang, X., Xie, H., & Delamater, P. L. (2015). Assessing the spatial accessibility of hospital care in Sichuan Province, China. *Geospatial Health*, *10*(2), 261–270. https://doi.org/10.4081/gh.2015.384.

Pan, X., Kwan, M. P., Yang, L., Zhou, S., Zuo, Z., & Wan, B. (2018). Evaluating the accessibility of healthcare facilities using an integrated catchment area approach. *International Journal of Environmental Research and Public Health*, *15*(9). https://doi.org/10.3390/ijerph15092051.

Piórkowski, A., Lupa, M., Bartoszewski, D., & Piorkowski, A. (2019). *The Comparison of Processing Efficiency of Spatial Data for PostGIS and MongoDB Databases*. https://doi.org/10.1007/978-3-030-19093-4_22.

Polo, G., Acosta, C. M., & Dias, R. A. (2013). Spatial accessibility to vaccination sites in a campaign against rabies in São Paulo city, Brazil. *Preventive Veterinary Medicine*, *111*(1–2), 10–16. https://doi.org/10.1016/j.prevetmed.2013.03.010.

Polzin, P., Borges, J., & Coelho, A. (2014). An extended kernel density two-step floating catchment area method to analyze access to health care. *Environ. Plann. B Plann. Des.*, *41*(4), 717–735. https://doi.org/10.1068/b120050p.

Preece, J., Rogers, Y., & Sharp, H. (2004). *Interaction design: Beyond Human Computer Interaction*. J. Wiley & Sons.

QGIS 2 Cookbook - Alex Mandel, Victor Olaya Ferrero, Anita Graser, Alexander Bruy - Google Books. (2016).

QGIS Python Programming Cookbook - Joel Lawhead - Google Books. (2015).

Radke, J., & Mu, L. (2000). Geographic Information Sciences A Journal of the Association of

Chinese Professionals in Geographic Information Systems Spatial Decompositions, Modeling and Mapping Service Regions to Predict Access to Social Programs Spatial Decompositions, Modeling and Mapping Service Regions to Predict Access to Social Programs. *Geomaahic Information Sciences*, *6*(2). https://doi.org/10.1080/10824000009480538.

Ranga, V., & Panda, P. (2014). Spatial access to in-patient health care in northern rural India. *Geospatial Health*, *8*(2), 545–556. https://doi.org/10.4081/gh.2014.44.

Rigby, J. E., & Gatrell, A. C. (2000). Spatial patterns in breast cancer incidence in north-west Lancashire. *Area*, *32*(1), 71–78. https://doi.org/10.1111/j.1475-4762.2000.tb00116.x.

Rocchini, D., Feilhauer, H., Delucchi, L., Bacaro, G., Cavallini, P., Foody, G. M., He, K. S., Nagendra, H., Porta, C., Ricotta, C., Schmidtlein, S., Spano, L. D., Wegmann, M., & Neteler, M. (2012). *Calculating landscape diversity with information-theory based indices: A GRASS GIS solution UAVforSAT-Operationalization of Vegetation Mapping through UAV-based Reference Data Acquisitions and Cloud-based Analysis of Earth Observation Data View project NO WASTE-ITAT2022 View project Calculating landscape diversity with information-theory based indices: A GRASS GIS solution*. https://doi.org/10.1016/j.ecoinf.2012.04.002.

Rocha, T. A. H., Da Silva, N. C., Amaral, P. V., Barbosa, A. C. Q., Rocha, J. V. M., Alvares, V., De Almeida, D. G., Thumé, E., Thomaz, E. B. A. F., De Sousa Queiroz, R. C., De Souza, M. R., Lein, A., Lopes, D. P., Staton, C. A., Vissoci, J. R. N., & Facchini, L. A. (2017). Addressing geographic access barriers to emergency care services: A national ecologic study of hospitals in Brazil. *International Journal for Equity in Health*, *16*(1), 1–10. https://doi.org/10.1186/s12939-017-0645-4.

Rutzinger, M., Höfle, B., & Stötter, J. (2006). *Object-based building detection based on airborne laser scanning data within GRASS GIS environment FloodTimeS: From Climate Change-Flood-Relationship to Flood Risk Time Series View project Mountains View project*. Retrieved February 26, 2020, from https://www.researchgate.net/publication/264085081.

Saeed, A. I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Quackenbush, J. (2003). TM4: A free, open-source system for microarray data management and analysis. *BioTechniques*, *34*(2), 374–378. https://doi.org/10.2144/03342mt01.

Sankepally, R. R., & Rajan, K. S. (2018). *IMPROVING PATH QUERY PERFORMANCE IN PGROUTING USING A MAP GENERALIZATION APPROACH*. https://doi.org/10.5194/isprs-archives-XLII-4-W8-191-2018.

Scavuzzo, C. M., Provensal, M. C., & Lanfri, M. (2017). *Generation of environmental information for public health applications using GRASS GIS EOXPOSURE (H2020 MSCA-RISE) View project How to coexist with fire in a flammable Córdoba (Argentina): an interdisciplinary approach View project*. https://doi.org/10.23919/RPIC.2017.8214336.

Schuurman, N., Bérubé, M., & Crooks, V. A. (2010). Measuring potential spatial access to primary health care physicians using a modified gravity model. *Canadian Geographer*, *54*(1), 29–45. https://doi.org/10.1111/j.1541-0064.2009.00301.x.

Scott, J., Larson, A., Jefferies, F., & Veenendaal, B. (2006). Small-area estimates of general practice workforce shortage in rural and remote Western Australia. *Australian Journal of Rural Health*, *14*(5), 209–213. https://doi.org/10.1111/j.1440-1584.2006.00811.x.

Sehra, S. S., Singh, J., & Rai, H. S. (2017). Assessing openstreetmap data using intrinsic quality indicators: An extension to the QGIS processing toolbox. *Future Internet*, *9*(2). https://doi.org/10.3390/fi9020015.

Shah, T. I., Bath, B., & Milosavljevic, S. (2015). Comparing geographical distribution of community-based physiotherapists and family physicians across Saskatchewan. *Canadian Geographer*, *59*(4), 461–473. https://doi.org/10.1111/cag.12224.

Shi, X., Alford-teaster, J., Onega, T., & Wang, D. (2012). *Spatial Access and Local Demand for Major Cancer Care Facilities in the United States Spatial Access and Local Demand for Major Cancer Care Facilities in the United States*. *October 2013*, 37–41. https://doi.org/10.1080/00045608.2012.657498.

Shin, H. & Ahn, E. (2018). Does the regional deprivation impact the spatial accessibility to dental care services? *Plos One*, *13*(9), e0203640. https://doi.org/10.1371/journal.pone.0203640.

Siegel, M., Koller, D., Vogt, V. & Sundmacher, L. (2016). Developing a composite index of spatial accessibility across different health care sectors: A German example. *Health Policy*, *120*(2), 205–212. https://doi.org/10.1016/j.healthpol.2016.01.001.

Singh, Y. P., Singh, A. K., & Singh, R. P. (2010). *Web GIS based Framework for Citizen Reporting on Collection of Solid Waste and Mapping in GIS for Allahabad City*. https://doi.org/10.18090/samriddhi.v8i1.11405.

Song, P., Zhu, Y., Mao, X., Li, Q., & An, L. (2013). Assessing Spatial Accessibility to Maternity Units in Shenzhen, China. *PLoS ONE*, *8*(7). https://doi.org/10.1371/journal.pone.0070227.

Steiniger, S., & Bocher, E. (2009). An overview on current free and open source desktop GIS developments. *International Journal of Geographical Information Science*, *23*(10), 1345–1370. https://doi.org/10.1080/13658810802634956.

Tan, Y., Song, Y., Song, Y., Wu, P., Cheng, J. C., Jeong Kim, M., & Wang, X. (2018). *Spatial and temporal variations of spatial population accessibility to public hospitals: A case study of rural-urban comparison GIScience & Remote Sensing Spatial and temporal variations of spatial population accessibility to public hospitals: a case study of rural-urban comparison Spatial and temporal variations of spatial population accessibility to public hospitals: a case study of rural-urban comparison*. https://doi.org/10.1080/15481603.2018.1446713.

Tang, J. H., Chiu, Y. H., Chiang, P. H., Su, M. D., & Chan, T. C. (2017). A flow-based statistical model integrating spatial and nonspatial dimensions to measure healthcare access. *Health and Place*, *47*, 126–138. https://doi.org/10.1016/j.healthplace.2017.08.006.

Tansley, G., Stewart, B., Zakariah, A., Boateng, E., Achena, C., Lewis, D., & Mock, C. (2016). Population-level Spatial Access to Prehospital Care by the National Ambulance Service in

Ghana. *Prehospital Emergency Care*, *20*(6), 768–775. https://doi.org/10.3109/10903127.2016.1164775.

Tao, Z., Yao, Z., Kong, H., Duan, F., & Li, G. (2018). Spatial accessibility to healthcare services in Shenzhen, China: improving the multi-modal two-step floating catchment area method by estimating travel time via online map APIs. *BMC Health Services Research*, *18*(1), 345. https://doi.org/10.1186/s12913-018-3132-8.

Taylor, P., & Wang, L. (2011). *Annals of GIS Analysing spatial accessibility to health care : a case study of access by different immigrant groups to primary care physicians in Toronto groups to primary care physicians in Toronto*. *April 2013*, 37–41.

Terry, M., Kay, M., & Lafreniere, B. (2010). Perceptions and practices of usability in the free/open source software (FOSS) community. *Conference on Human Factors in Computing Systems - Proceedings*, *2*, 999–1008. https://doi.org/10.1145/1753326.1753476.

The Open Source Definition | Open Source Initiative. Retrieved August 2, 2020, from https://opensource.org/osd.

Tomintz, M.N., Clarke, G.P & Rigby, J.E. The geography of smoking in Leeds: estimating individual smoking rates and the implications for the location of stop smoking services. *Area*. 2008; 40(3):341-53

V.Bhanumurthy, Bothale, V. M., Kumar, B., Urkude, N., & Shukla, R. (2015). Route Analysis for Decesion Support System in Emergency Management Through Gis Technologies. *International Journal of Advanced Engineering and Global Technology*, *3*(2), 345–350. www.ijaegt.com.

Vadrevu, L., & Kanjilal, B. (2016). Measuring spatial equity and access to maternal health services using enhanced two step floating catchment area method (E2SFCA) - A case study of the Indian Sundarbans. *International Journal for Equity in Health*, *15*(1), 1–12. https://doi.org/10.1186/s12939-016-0376-y.

Vázquez-Rodríguez, R., Carlos Pérez-Risquet & Carlos Torres-Cantero, J. (2013). *A Novel*

Visual Data Mining Module for the Geographical Information System gvSIG Um Novo Módulo de Mineração VISUAL de Dados para o Sistema de Informação Geográfica gvSIG. https://doi.org/10.11137/2013_1_98_111.

Von Hippel, E., & Von Krogh, G. (2003). *open Source Software and the "Private-Collective" Innovation Model: Issues for Organization Science*.

Walters, S., Maringe, C., Butler, J., Rachet, B., Barrett-Lee, P., Bergh, J., Boyages, J., Christiansen, P., Lee, M., Wärnberg, F., Allemani, C., Engholm, G., Fornander, T., Gjerstorff, M. L., Johannesen, T. B., Lawrence, G., McGahan, C. E., Middleton, R., Steward, J. Coleman, M. P. (2013). Breast cancer survival and stage at diagnosis in Australia, Canada, Denmark, Norway, Sweden and the UK, 2000-2007: A population-based study. *British Journal of Cancer*, *108*(5), 1195–1208. https://doi.org/10.1038/bjc.2013.6.

Wan, N., Zhan, F. B., Lu, Y., & Tiefenbacher, J. P. (2012). Access to healthcare and disparities in colorectal cancer survival in Texas. *Health and Place*, *18*(2), 321–329. https://doi.org/10.1016/j.healthplace.2011.10.007.

Wan, N., Zhan, F. B., Zou, B., & Wilson, J. G. (2013). Spatial Access to Health Care Services and Disparities in Colorectal Cancer Stage at Diagnosis in Texas. *The Professional Geographer*, *65*(3), 527–541. https://doi.org/10.1080/00330124.2012.700502.

Wan, N., Zou, B., & Sternberg, T. (2012). A three-step floating catchment area method for analyzing spatial access to health services. *International Journal of Geographical Information Science*, *26*(6), 1073–1089. https://doi.org/10.1080/13658816.2011.624987.

Wang, Fahui. (2000). Modeling Commuting Patterns in Chicago in a GIS Environment: A Job Accessibility Perspective. *Professional Geographer*, *52*(1), 120–133. https://doi.org/10.1111/0033-0124.00210.

Wang, Fahui. (2018). Inverted Two-Step Floating Catchment Area Method for Measuring Facility Crowdedness. *Professional Geographer*, *70*(2), 251–260. https://doi.org/10.1080/00330124.2017.1365308.

Wang, Fahui, & Luo, W. (2005). Assessing spatial and nonspatial factors for healthcare access: Towards an integrated approach to defining health professional shortage areas. *Health and Place*, *11*(2), 131–146. https://doi.org/10.1016/j.healthplace.2004.02.003.

Wang, Fahui, McLafferty, S., Escamilla, V., & Luo, L. (2008). Late-Stage Breast Cancer Diagnosis and Health Care Access in Illinois. *The Professional Geographer*, *60*(1), 54–69. https://doi.org/10.1080/00330120701724087.

Wang, Fuyuan, & Wang, K. (2018). Measuring spatial accessibility to ecological recreation spaces in the Pearl River delta region: an improved two-step floating catchment area method. *Journal of Spatial Science*, *63*(2), 279–295. https://doi.org/10.1080/14498596.2018.1488633.

Wang, L., & Ramroop, S. (2018). Geographic disparities in accessing community pharmacies among vulnerable populations in the Greater Toronto Area. *Canadian Journal of Public Health*, *109*(5–6), 821–832. https://doi.org/10.17269/s41997-018-0110-1.

Wang, X., & Pan, J. (2016). Assessing the disparity in spatial access to hospital care in ethnic minority region in Sichuan Province, China. *BMC Health Services Research*, *16*(1), 1–11. https://doi.org/10.1186/s12913-016-1643-8.

Wiki, J., Kingham, S., & Campbell, M. (2019). Accessibility to food retailers and socio-economic deprivation in urban New Zealand. *New Zealand Geographer*, *75*(1), 3–11. https://doi.org/10.1111/nzg.12201.

Wu, H., Liu, L., Yu, Y., & Peng, Z. (2018). Evaluation and Planning of Urban Green Space Distribution Based on Mobile Phone Data and Two-Step Floating Catchment Area Method. *Sustainability*, *10*(1), 214. https://doi.org/10.3390/su10010214.

Xia, T., Song, X., Zhang, H., Song, X., Kanasugi, H., & Shibasaki, R. (2019). Measuring spatio-temporal accessibility to emergency medical services through big GPS data. *Health and Place*, *56*, 53–62. https://doi.org/10.1016/j.healthplace.2019.01.012.

Xie, An, Zheng, & Li. (2018). *Healthy aging with parks: Association between park accessibility and the health status of older adults in urban China*. https://doi.org/10.1016/j.scs.2018.09.010.

Xing, L., Liu, Y., & Liu, X. (2018). Measuring spatial disparity in accessibility with a multi-mode method based on park green spaces classification in Wuhan, China. *Applied Geography*, *94*(March), 251–261. https://doi.org/10.1016/j.apgeog.2018.03.014.

Xu, Y., Fu, C., Onega, T., Shi, X., & Wang, F. (2017). Disparities in Geographic Accessibility of National Cancer Institute Cancer Centers in the United States. *Journal of Medical Systems*, *41*(12), 1–11. https://doi.org/10.1007/s10916-017-0850-0.

Ye, C., Zhu, Y., Yang, J., & Fu, Q. (2018). Spatial equity in accessing secondary education: Evidence from a gravity-based model. *Canadian Geographer*, *62*(4), 452–469. https://doi.org/10.1111/cag.12482.

Yin, P. (2018). Does Better Spatial Access Lead to Better Use of Prenatal Care? A Population Study in Georgia. *Professional Geographer*, *70*(3), 363–373. https://doi.org/10.1080/00330124.2017.1385402.

Yoshida, D., Song, X., & Raghavan, V. (2010). Development of track log and point of interest management system using free and open source software. *Applied Geomatics*, *2*(3), 123–135. https://doi.org/10.1007/s12518-010-0028-7.

Yu, D., Tan, H., & Ruan, Y. (2012). An improved two-step floating catchment area method for supporting district building energy planning: A case study of Yongding County city, China. *Applied Energy*, *95*, 156–163. https://doi.org/10.1016/j.apenergy.2012.02.036.

Zahnd, W. E., McLafferty, S. L., Sherman, R. L., Klonoff-Cohen, H., Farner, S., & Rosenblatt, K. A. (2019). Spatial Accessibility to Mammography Services in the Lower Mississippi Delta Region States. *The Journal of Rural Health*, *35*(4), 550–559. https://doi.org/10.1111/jrh.12349.

Appendices

Appendix A Table of papers relating to cancer

Author	Date	Method	Synopsis
Wang et al	2008	2SFCA	Investigates the link between late stage breast cancer diagnosis, access to primary healthcare and mammography screening services in Illinois, US. Using 2SFCA the study showed that poor accessibility to primary healthcare increased the late stage breast cancer diagnosis, particularly noting that those living in a high concentration of disadvantaged people are at the most significant risk. There was no significant link between mammography screening accessibility and late stage breast cancer diagnosis in this study. The authors point out the limitations with the data they used but, this study shows important results for healthcare policy makers.
Gu et al	2010	Modified 2SFCA	Suggests a new framework to measure accessibility, building on the 2SFCA method the authors add a distance factor and a Huff-based competitive model. The authors conduct a case study on breast cancer screening services in Alberta, Canada and the results show that it is beneficial when planning where to locate preventative healthcare facilities.
McLafferty et al	2011	2SFCA	Investigated the rural-urban inequalities in late stage breast cancer diagnosis for both the overall population and the African-American population, in two time periods 1988-1992 and 1998-2002. They did this by using several techniques including 2SFCA for the spatial aspect of the paper. The results showed that there was no significant rural-urban inequality for the overall population, there was a decline in late stage breast cancer diagnosis between the two-time periods with bigger decreases in higher population areas. There was a large disparity between the overall population and the African-American population in both time periods.

Lian et al	2012	2SFCA	Evaluated several different ways of evaluating the link between accessibility to mammography services and late stage cancer. Two 2SFCA models were used and they both showed an increase in odds of late stage breast cancer. The methods evaluated were 2SFCA, travel time and service density.
Wan et al	2012	E2SFCA	Using E2SFCA and aspatial methods to investigate the links between accessibility and colorectal cancer in the Texas, US region. The study used three weights for the different oncologists in each catchment area. The study showed that accessibility to oncologists did make a difference in rural areas but not in urban areas. The aspatial data showed that disadvantaged groups had more chance of dying from colorectal cancer.
Wang	2012	2SFCA, KD2SFCA	Evaluates current methodology from the gravity model through 2SFCA and includes KD2SFCA, and discusses different decay functions. The paper discusses the need for greater utilisation of spatial and aspatial data and discusses the implications it has on mortality rates using cancer in the USA as an example. The paper suggests a framework should be implemented and suggests using optimisation techniques.
Wan et al	2012	E2SFCA	Introduces the idea of using a spatial access ratio derived from the E2SFCA spatial access index (SPAI). A case study into the accessibility of colorectal cancer facilities in Texas, US is used to trial the proposed methods. The study used a sensitivity analysis to show that the SPAI value fluctuates with changes in the impedance coefficient whereas the SPAR stays stable throughout. The authors note that if the impedance is from some basis and not an arbitrary input then the SPAI value would be valuable and could show more detailed results.
Wan et al	2012	E2SFCA	Using E2SFCA and a factor analysis to assess the effect of

			spatial and aspatial accessibility to oncologists in Texas, US. The study examines the effect of accessibility with regard to colorectal cancer survival, the E2SFCA method used a Gaussian function and a three-step catchment area. Cox proportional hazard regression was used to analyse the association between healthcare accessibility and colorectal cancer survival. The factor analysis and principal components analysis was completed and nine aspatial data were included in the study, including household income, home valuation and high school education. The results showed that several minority groups were more likely to die from colorectal cancer and that spatial access to oncologists had a significant effect on colorectal survival in non-urban areas but not in urban areas themselves.
Shi et al	2012	2SFCA	Using 2SFCA to evaluate accessibility to cancer care facilities in the USA. The study suggests that the travel impedance threshold is the crucial factor when conducting a 2SFCA calculation as this is a special case of the gravity model. It also suggests that different choices in distance decay should not affect the overall pattern when used. This is a very large area for a 2SFCA study and it showed that there were two main regions with lower accessibility; the east and the west of the US, the west had deeper pockets of poor accessibility whereas the east had larger areas with slightly lower accessibility ratings.

Table 1

Appendix B Table of 2SFCA papers

Author	Year	Method	Subject area	aspatial data
Luo & Wang	2003	2SFCA	Primary healthcare physician accessibility in Chicago	N
Wang & Luo	2004	2SFCA	Primary healthcare accessibility in Illinois combining spatial and aspatial data	Y
Luo	2004	2SFCA	Primary healthcare physician accessibility in Northern Illinois	N
Wang	2007	2SFCA	Access to healthcare for Chinese immigrants in Toronto using spatial and aspatial data	Y
Wang et al	2008	2SFCA	Late stage breast cancer diagnosis and access to primary healthcare in Illinois	N
Langford et al	2008	2SFCA & Dasymetric 2SFCA	Using postcode database data and raster maps to balance the effects of heavily populated areas and give clearer results	N
McGrail & Humphreys	2009	2SFCA	Accessibility to primary healthcare in Victoria, Australia	N
Luo & Qi	2009	2SFCA & E2SFCA	Primary healthcare physicians in Northern Illinois	N
McGrail & Humphreys	2009	Modified 2SFCA	Access to healthcare in Victoria, Australia using spatial and aspatial data	Y
Gu et al	2010	Modified 2SFCA	Breast cancer screening services accessibility in Alberta, Canada	N
Wang	2011	2SFCA	Accessibility to primary healthcare for different types of immigrant in Toronto using spatial and aspatial data	Y
Ngui & Apparicio	2011	2SFCA & Optimised 2SFCA	Accessibility to healthcare in clinics in Montreal	N
McLafferty et al	2011	2SFCA	Investigated the inequalities between rural and urban populations and the African-American	Y

			and overall population at different times for late stage breast cancer diagnosis	
Lian et al	2012	2SFCA	Investigated the links between accessibility to mammography services and late stage breast cancer diagnosis	N
Wan et al	2012	3SFCA	Access to primary healthcare physicians in the central Texas region	N
Wan et al	2012	E2SFCA	Used spatial and aspatial data to investigate the link between colorectal cancer and healthcare accessibility in Texas	Y
Luo & Whippo	2012	V2SFCA	This method uses small increments in catchment size until a predetermined supply to demand ratio is met. This method attempts to address the arbitrary nature of catchment size selection and make the method more effective over larger and more rural areas.	N
McGrail	2012	2SFCA	Investigates the effects of variable catchment sizes and distance decay functions in rural Victoria, Australia for healthcare accessibility	N
Ngui & Vanasse	2012	Modified 2SFCA	Mental healthcare accessibility in Montreal	N
Bisonnette et al	2012	3SFCA	Neighbourhood level access to primary healthcare in Ontario, Canada	N
Bell et al	2012	3SFCA	Primary healthcare accessibility	N
Wang	2012	2SFCA & KD2SFCA	Uses spatial and aspatial data to investigate the effects of accessibility on cancer mortality rates in the USA	Y
Wan et al	2012	E2SFCA	Colorectal cancer facility accessibility in Texas	N
Wan et al	2012	E2SFCA	Uses spatial and aspatial data to investigate accessibility to Oncology in Texas	Y
Shi et al	2012	2SFCA	Access to cancer care facilities in the USA	N
Delameter	2013	E2SFCA, 3SFCA &	Acute care hospitals in Michigan, accessibility	N

		M2SFCA		
Dewulf et al	2013	2SFCA & E2SFCA	Used different methods to investigate accessibility to primary healthcare in Belgium	N
Song et al	2013	2SFCA	Access to maternity units in Shenzen, China	N
Hu et al	2013	E2SFCA	Access to healthcare in Donghai County and compare the urban and rural results	N
Mao & Nekorchuk	2013	2SFCA & a modified Multi-transport E2SFCA	Access to primary healthcare in Florida	N
McGrail & Humphreys	2014	V2SFCA	Access to healthcare across the whole of Australia	N
Polzin et al	2014	KD2SFCA	Aim to improve the standard KD2SFCA method and investigate access to healthcare facilities in Portugal	N
Ranga & Panda	2014	3SFCA	Access to inpatient care in rural Northern India	N
Luo	2014	Modified 2SFCA	Access to healthcare in Springfield, MO using a huff-model based population selection technique	N
Shah et al	2015	3SFCA	Compared the accessibility of family physicians and physiotherapists in Saskatchewan, Canada	N
Jamtsho et al	2015	Nearest neighbour modified 2SFCA (NN-M2SFCA)	This modification is intended to be used over large areas; this study investigated access to primary healthcare in Bhutan.	N
Barona & Blaschke	2015	2SFCA	Used spatial and aspatial data to investigate the health centre accessibility in Quito, Ecuador	Y

Pan et al	2015	E2SFCA	Access to hospitals in Sichuan Province, China	N
Lin et al	2016	Modified 2SFCA	Investigating accessibility to automated external defibrillator accessibility in Kaohsiung, Taiwan using spatial and aspatial data	Y
Jones et al	2016	3SFCA	Using spatial and aspatial data to compare the accessibility of dental services and primary healthcare in Saskatoon, Saskatchewan	Y
Donohoe et al	2016	2SFCA	Compared different types of accessibility measures in the Appalachia region of the USA	N
Cheng et al	2016	KD2SFCA	Access to high level hospitals in Shenzhen, China	N
Gao et al	2016	E2SFCA	Investigated access to healthcare services for pregnant women in France	N
Kanuganti et al	2016	E2SFCA	Healthcare accessibility in Rajasthan, India	N
Wang & Pan	2016	E2SFCA	Investigated the inequity of hospital accessibility between ethnic minorities and the overall population in Sichuan Province, China	Y
Cao et al	2016	E2SFCA	Investigation into end stage renal disease in Midwestern USA. The study analyses spatial and aspatial data to see if variations in end stage renal disease occur between White, Black and Native American and rural/urban populations	Y
Langford et al	2016	Multi-Modal 2SFCA (MM-2SFCA)	Access to primary healthcare in South Wales using two different types of transportation network	N
Luo et al	2016	E3SFCA	Access to primary healthcare in Springfield, MO	N
Vadrevu & Kanijilal	2016	E2SFCA	Accessibility to healthcare for pregnant women in the Sundarbans region of India	N

Table 1

Appendix C Documentation

C.1 End User Documentation

Tool layout

This document will explain the layout of the tool and provide an example to follow when you use it for the first time. Some of the functions may not be applicable to you and these will be explained as and when they arise. The following section will expand on this and detail the method for visualising results in QGIS.

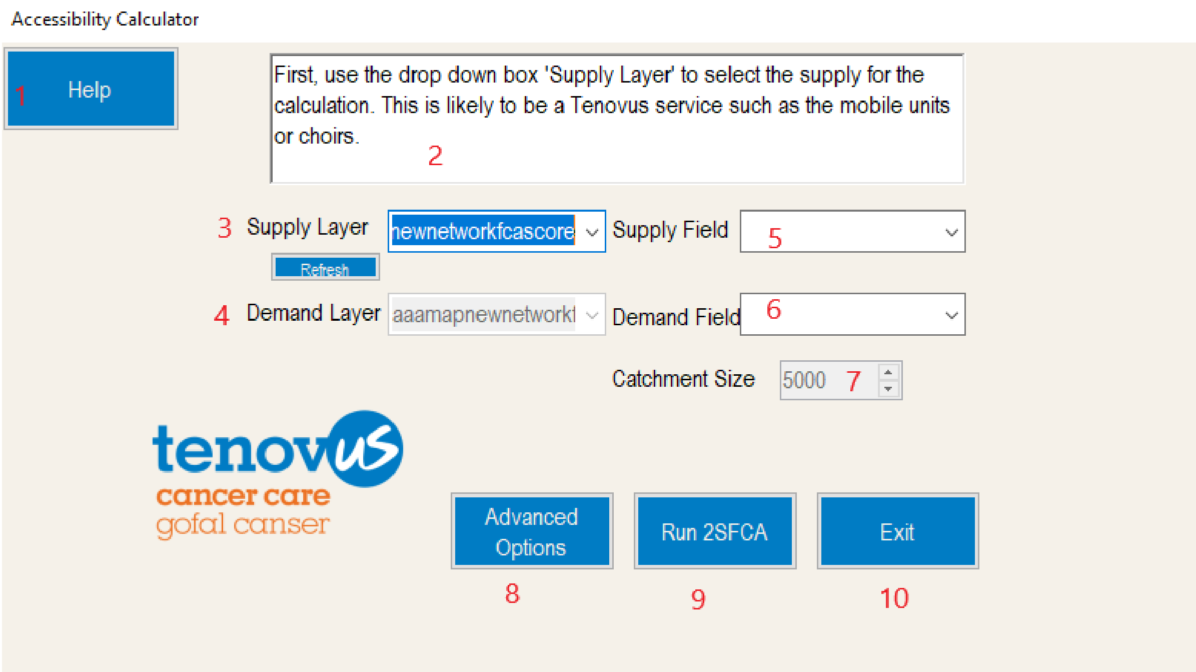


Figure 1: 2SFCA tool (main page)

Figure 1 shows the main page of the 2SFCA tool which is where most of the inputs are added and program starts. The numbers above relate to the functions;

1. The help button opens this document. It can be used if you get stuck in the future and require detailed assistance

2. This is the help screen and it will guide you through the required steps form the start, if you become stuck at any point this is a good place to get a hint on the next task to be completed
3. Several inputs are required for the program to run and this is the first, it relates to the supply layer, the supply is what people may need to access (mobile unit). This is likely to be a Tenovus service such as mobile units or choirs it could also be used to look at other data like the NHS chemotherapy locations or NHS lymphoedema locations
4. This is the demand layer which is similar to the supply layer but it is likely to represent the public and will probably be the census data (total population) or some data from the Welsh Index of Multiple Deprivation (WIMD) (Cancer rate)
5. The supply field is the particular part of the data that you are interested in, this may be the supply value (number of beds/chairs multiplied by the number of hours the site is open) or the number of beds within a unit
6. The demand value looks at the specifics of the layer, there will probably be a demand value to select or it may be total population or over 60s population
7. The catchment area size is the distance people will travel for the service, using Tenovus data it was possible to see that most of the Tenovus users travel 35km to utilise the services offered. The catchment size is very important to the calculation, this will vary depending upon what service is being assessed. If it is a very niche or specialised service (chemotherapy) the catchment will be larger than if it is a very popular or often used service (dentist, optician)
8. This button will open the advanced options screen
9. This button will use all of the inputs added above and run the 2SFCA calculation
10. This is the exit button

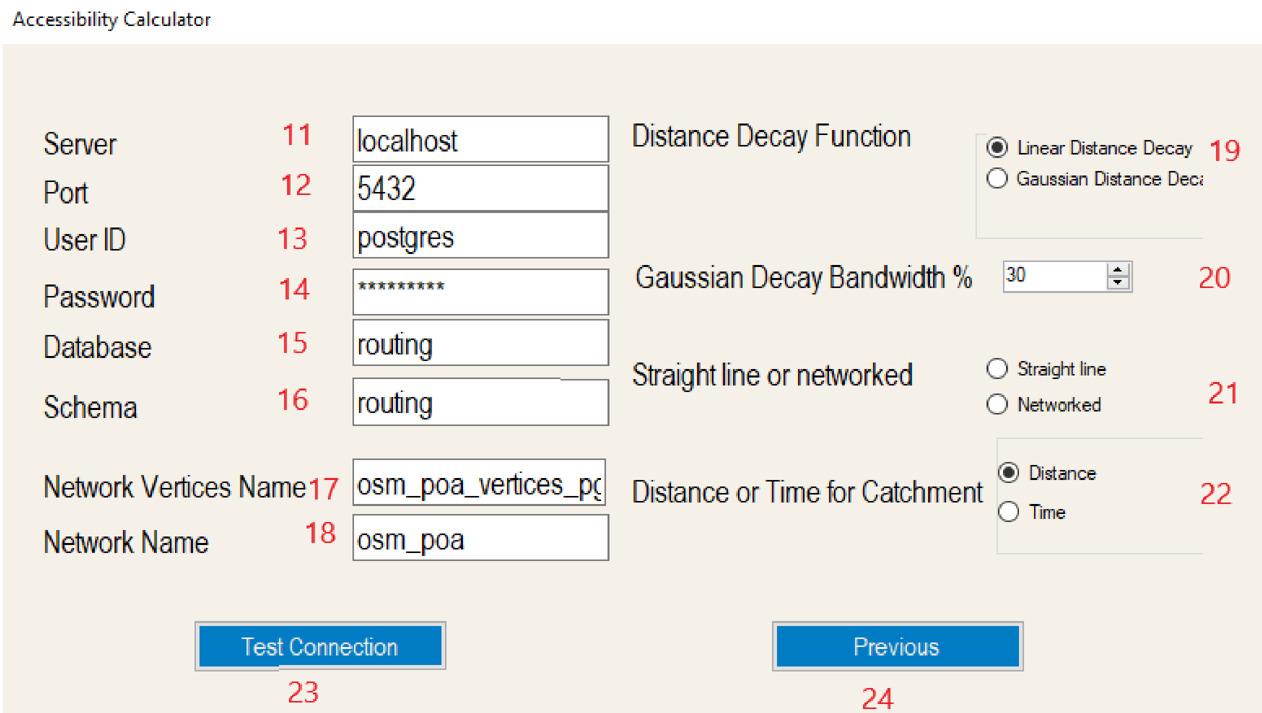


Figure 2: 2SFCA tool (advanced options page)

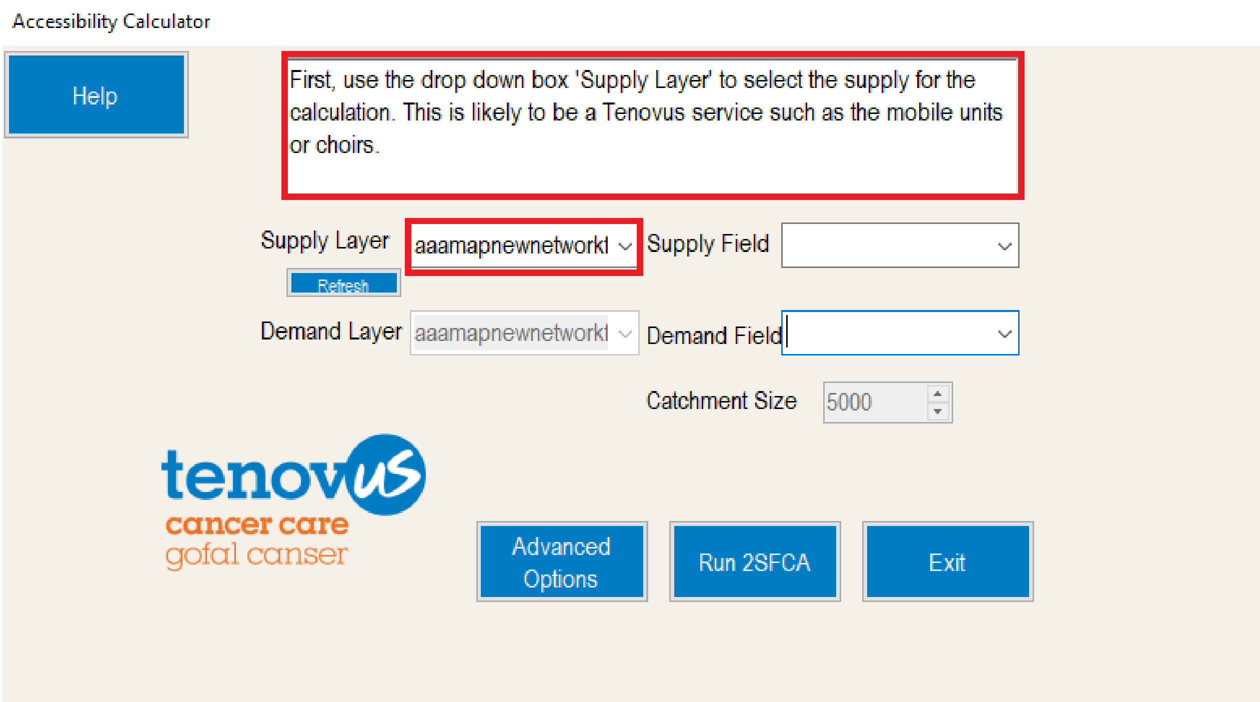
Numbers 11-18 are an input for the server and should have been set up by the super user using part 2 of the documentation

11. This is the server that the tool is connecting with to retrieve the information
12. This is the port that the tool is connecting with
13. This is the user ID
14. This is the password for the connection to the database
15. This is the database that the tool is connecting with
16. This is the schema on the database
17. To be able to get accurate times a routable road network is required and this is the network vertices name
18. This is the network name
19. There are several types of distance decay function which can be used, as standard the linear distance decay is setup but this can be changed to a Gaussian distance decay or no distance decay. See section 3 for more information on the different distance decay functions
20. This is the size of the curve only used with the Gaussian distance decay

21. It is possible to run the tool by using the road network to estimate travel time (set up as standard) or without this and a straight-line is connected between the supply and demand point (less accurate)
22. It is possible to use distance in metres (standard) or time in minutes for the catchment size. It may be easier to say that people travel for 30 minutes to access a service rather than 25,000 metres
23. This is the test connection button and allows a user to ensure they are connected to the server (a pop-up with connection good will appear)
24. This button returns to the main screen

Using the tool

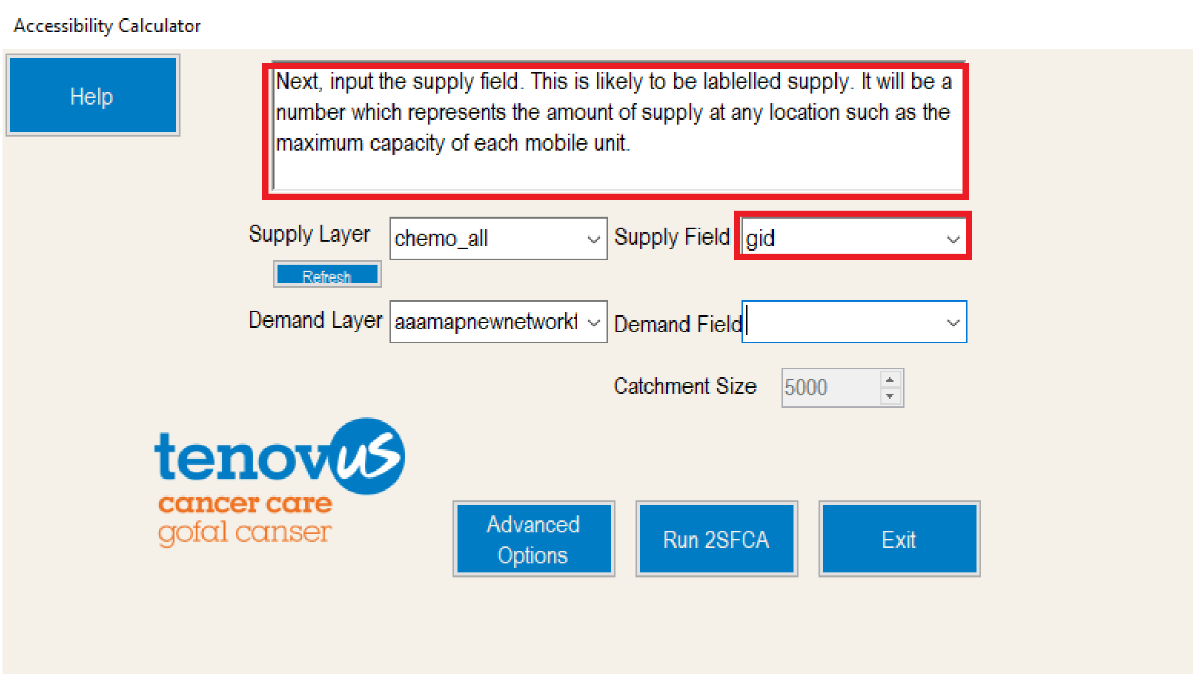
This section will run through an example of how to use the tool. By following the 6 below the 2SFCA method will run and a results table will be created in the database for visualisation.



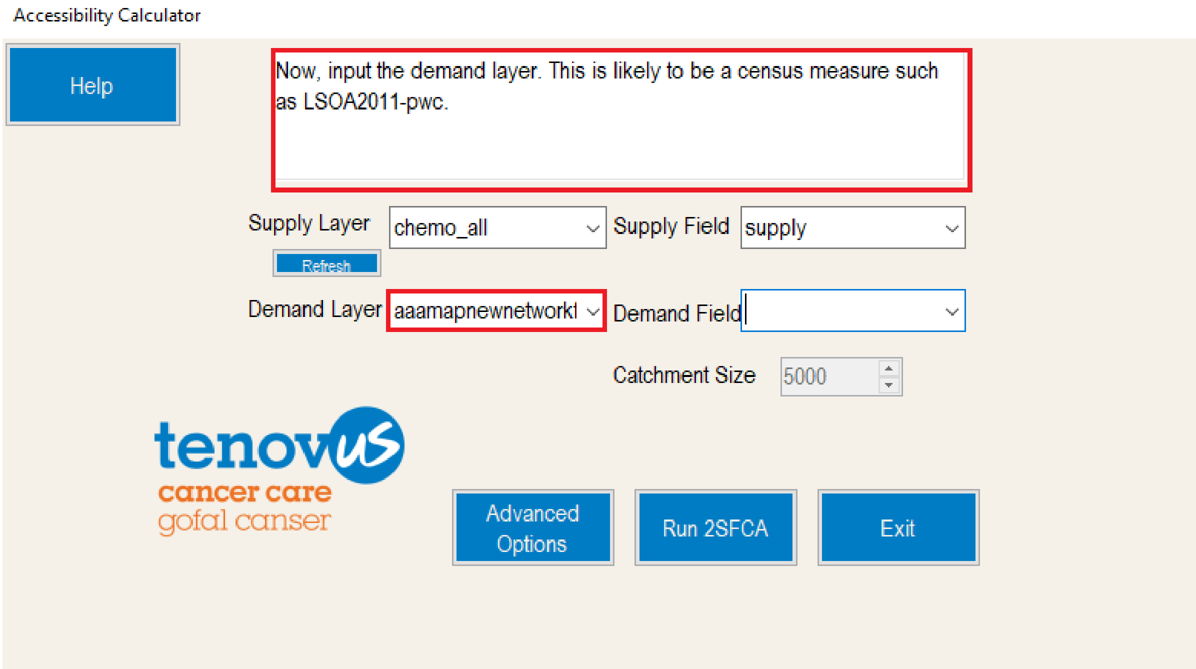
The first thing that needs to be entered is the supply layer. This refers to the table within the database which holds all of the information on a supply of services. In this example chemo_all is selected and this dataset holds all of the data regarding the different sites in Wales which provide chemotherapy services. Most of this data is not used in the front end and is required for the program to run (see below).

	gid [PK] serial	location f character varying(254)	unit character varying(254)	nurses_wte numeric	open_weekl numeric	beds_chair numeric(10,0)	supply numeric	capacity character varying(254)	postcode character varying(254)	x_east numeric(10,0)	y_north numeric(10,0)	geom geometry(Point,27700)
1	1	Ysbyty Maelor Wrecsam	no	9.15000000000000	42.5000000000000	11	467.500000000000	30	LL13 7TD	332375	350356	0101000020346C0000000000005C491441000000050621541
2	2	Ysbyty Glan Clwyd	no	9.85999999999999	42.5000000000000	21	892.500000000000	40	LL18 5UJ	300309	376019	0101000020346C000000000005454124100000004CF31641
3	3	Ysbyty Gwynedd	no	7.31000000000000	42.5000000000000	20	850.000000000000	30	LL57 2PW	255857	370173	0101000020346C0000000000883B0F4100000000F4971641
4	4	Singleton Hospital	no	45.3900000000001	50.0000000000000	36	1800.00000000000	55 (avg per day)	SA2 8QA	262626	191910	0101000020346C000000000008807104100000000306D0741
5	5	Royal Gwent Hospital	no	9.00000000000000	40.0000000000000	26	1040.00000000000		NP20 2UB	330932	187243	0101000020346C00000000000D032144100000000058DB0641
6	6	Nevill Hall Hospital	no	5.00000000000000	32.0000000000000	11	352.000000000000		NP7 7EG	328833	214506	0101000020346C000000000000412144100000000502F0A41
7	7	Bronglais General Hospital	no	2.76000000000000	40.0000000000000	6	240.000000000000		SY23 1ER	259208	281805	0101000020346C00000000000040A40F410000000034331141
8	8	Glangwili General Hospital	no	4.80000000000000	40.0000000000000	8	320.000000000000		SA31 2AF	242792	221240	0101000020346C00000000000040A30D4100000000C0010B41
9	9	Prince Philip Hospital	no	3.20000000000000	40.0000000000000	6	240.000000000000		SA14 8QF	252459	201367	0101000020346C0000000000058D10E4100000000B8940841
10	10	Withybush General Hospital	no	4.40000000000000	40.0000000000000	8	320.000000000000		SA61 2PZ	195710	216835	0101000020346C00000000000F0E307410000000018780A41
11	11	Chemo inpatient unit Velindre Cancer Centre	no	4.00000000000000	45.0000000000000	14	630.000000000000	6	CF14 2TL	314817	180435	0101000020346C000000000004371341000000098060641
12	12	Chemo day unit Velindre	no	6.00000000000000	45.0000000000000	12	540.000000000000	30	CF14 2TL	314817	180435	0101000020346C000000000004371341000000098060641
13	13	Rhosyn day unit Velindre	no	5.00000000000000	42.5000000000000	9	382.500000000000	25	CF14 2TL	314817	180435	0101000020346C000000000004371341000000098060641
14	14	Prince Charles Hospital	no	5.00000000000000	40.0000000000000	10	400.000000000000	24	CF47 9DT	304481	208161	0101000020346C00000000000849512410000000008690941
15	15	Royal Glamorgan Hospital	no	2.00000000000000	30.0000000000000	3	90.0000000000000	12	CF72 8XR	303649	184110	0101000020346C00000000000848812410000000070790641
16	16	Nevill Hall Hospital	no	5.00000000000000	40.0000000000000	10	400.000000000000	24	NP7 7EG	328833	214506	0101000020346C000000000000412144100000000502F0A41
17	17	Princess of Wales Hospital	no	2.00000000000000	25.0000000000000	3	75.0000000000000	12	CF31 1RQ	291047	181033	0101000020346C0000000000009CC311410000000048190641
18	18	Ysbyty Ystrad Fawr	no	2.00000000000000	15.0000000000000	4	60.0000000000000	6	CF82 7EP	314564	193444	0101000020346C00000000000103313410000000209D0741
19	19	St Davids	no	2.00000000000000	27.5000000000000	4	110.000000000000	12	CF11 9XB	317297	176583	0101000020346C0000000000000C45D134100000000388E0541
20	20	Ysbyty Aneurin	no	2.00000000000000	15.0000000000000	3	45.0000000000000	6	NP23 6GL	317138	209045	0101000020346C000000000000485B1341000000000A8840941
21	21	University Hospital Wales	no	24.0000000000000	50.0000000000000	27	1350.00000000000		CF14 4XW	317551	179319	0101000020346C0000000000000BC6113410000000B8E30541
22	22	University Hospital Wales Llandough	no	2.00000000000000	20.0000000000000	4	80.0000000000000		cf64 2xx	316526	172981	0101000020346C0000000000000B851134100000000A81D0541
23	23	Ward C4 Neuro Day Unit	no	5.00000000000000	45.0000000000000	7	315.000000000000	12	CF14 4XW	317551	179319	0101000020346C0000000000000BC6113410000000B8E30541
24	24	Cwmbran	yes		8.00000000000000	7	56.0000000000000		NP44 3YS	329434	194301	0101000020346C00000000000681B1441000000000E8B70741
25	25	Nantgarw	yes		8.00000000000000	7	56.0000000000000		CF15 7QX	311723	185868	0101000020346C0000000000000AC06134100000000060B00641
26	26	Swansea	yes		8.00000000000000	7	56.0000000000000		SA1 3ST	265757	192701	0101000020346C00000000000743810410000000000E8850741

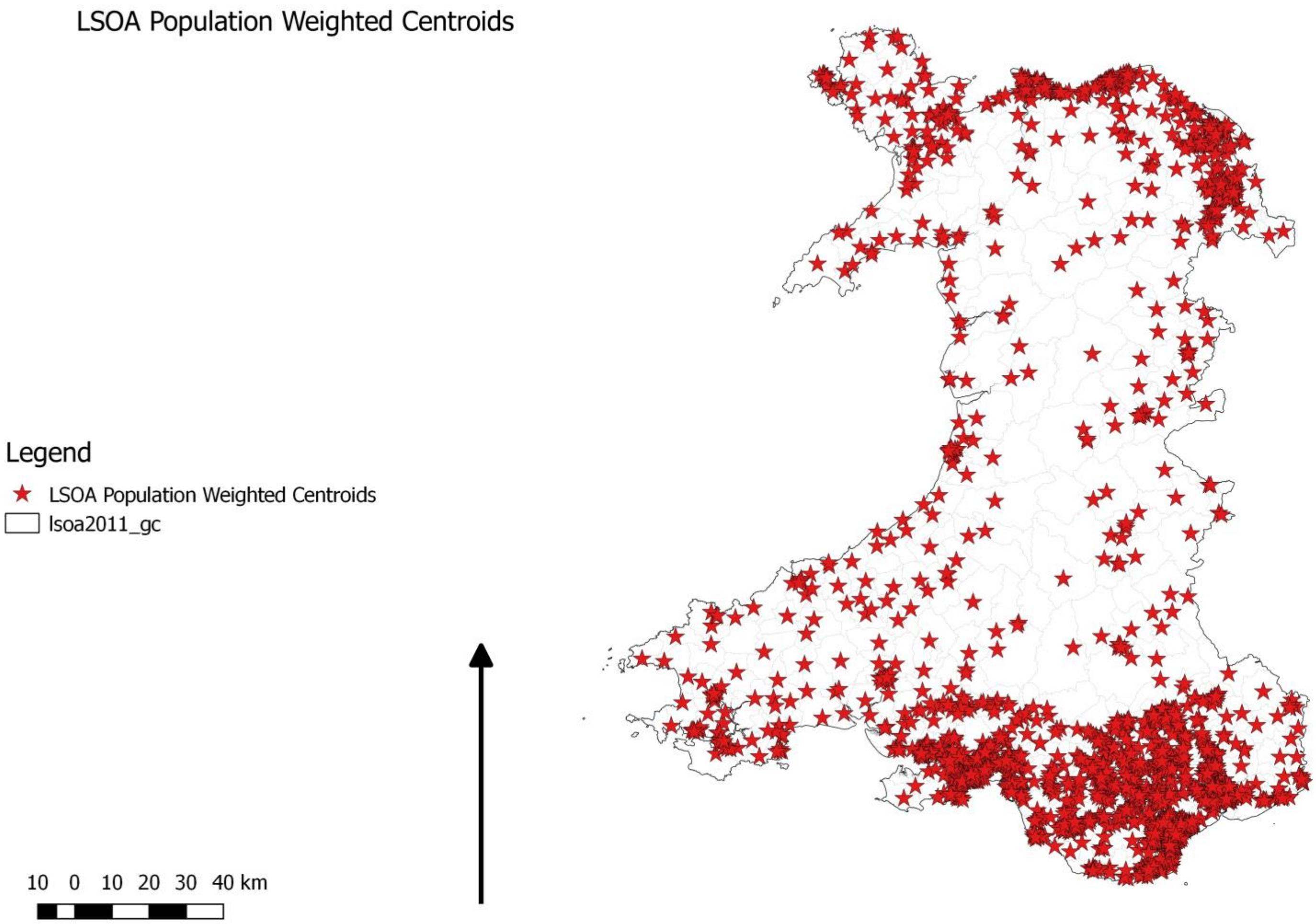
The database holds lots of information and can show whether the location is a hospital or a Tenovus unit. The interesting columns for entry would be nurses_wte which is the whole time equivalent of the nurses working at the location, capacity which is the number of patients that can be treated or the open_weekly column which shows the number of hours each location is open per week.



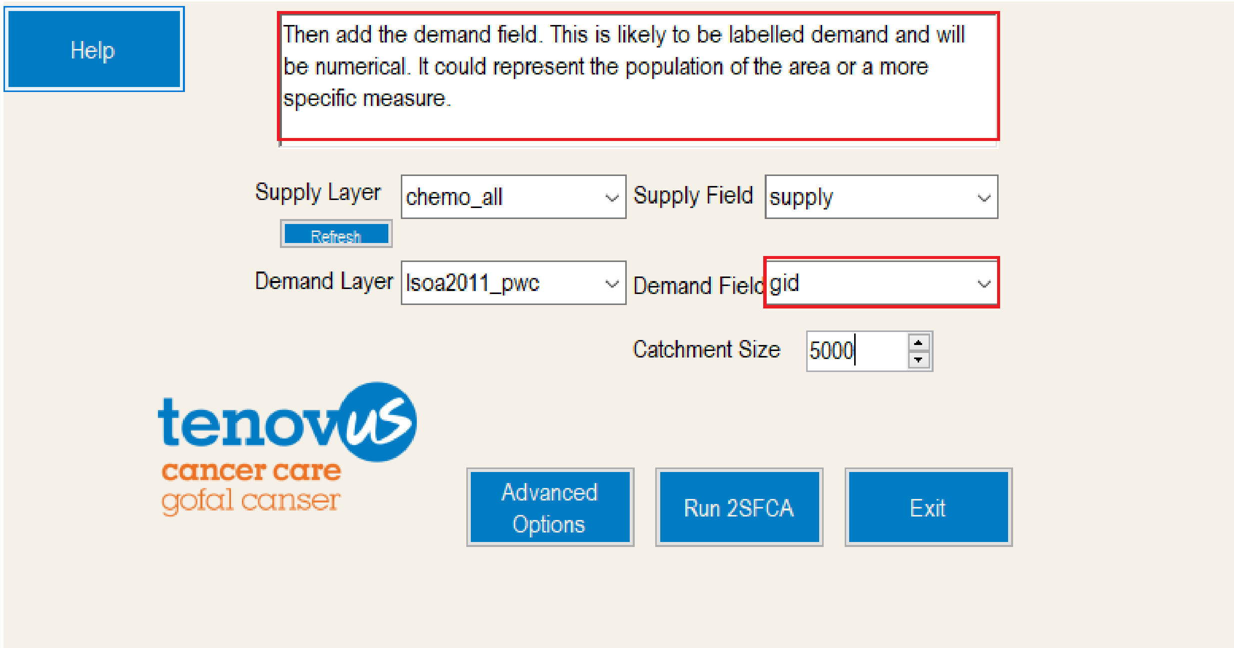
Next, the column from the database needs to be selected, the supply column is being used in this example and that takes the number of beds or chairs at each location and multiplies it by the opening hours to give a good overview of the amount of supply each location provides.



The demand layer is similar to the supply layer but it is used to represent the people who use the service so this could be defined by the current cancer rate of each location, the average age of the population at each location or just the total population at each demand point. The data used here is at LSOA level which is a way of categorising households for census data. The below map shows the different LSOAs in Wales and each one has a minimum of 1000 people with a mean of 1500.

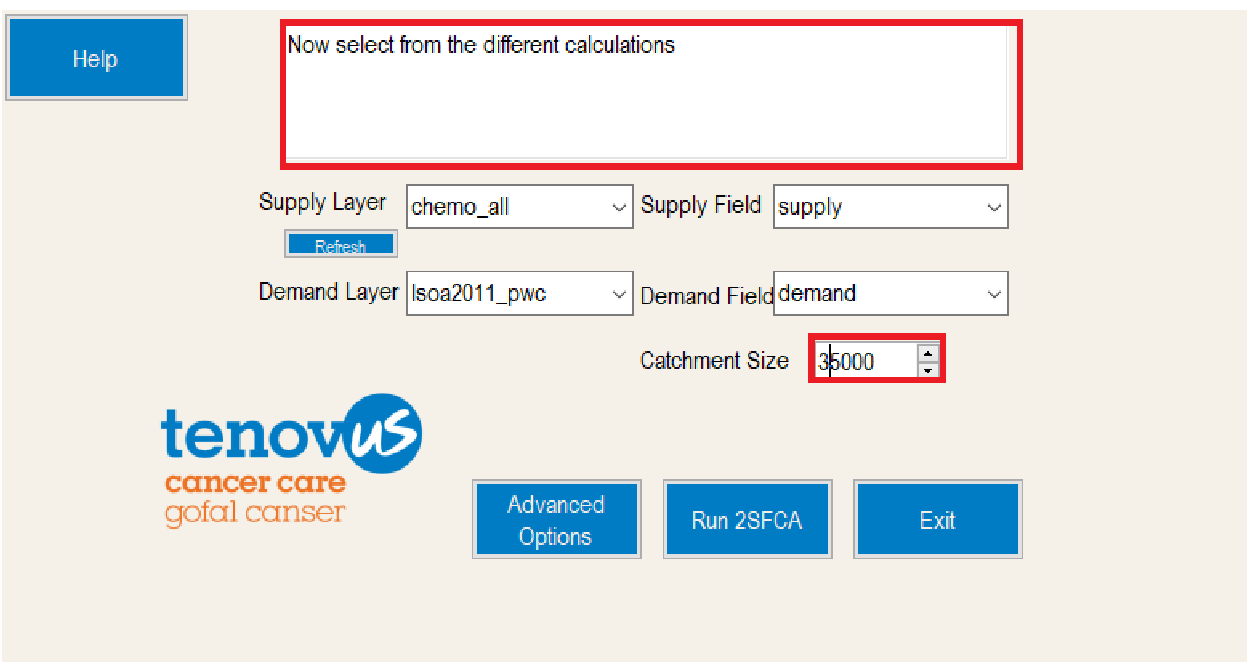

LSOA Population Weighted Centroids
Legend
LSOA Population Weighted Centroids
lsoa2011_gc
10 0 10 20 30 40 km

Accessibility Calculator

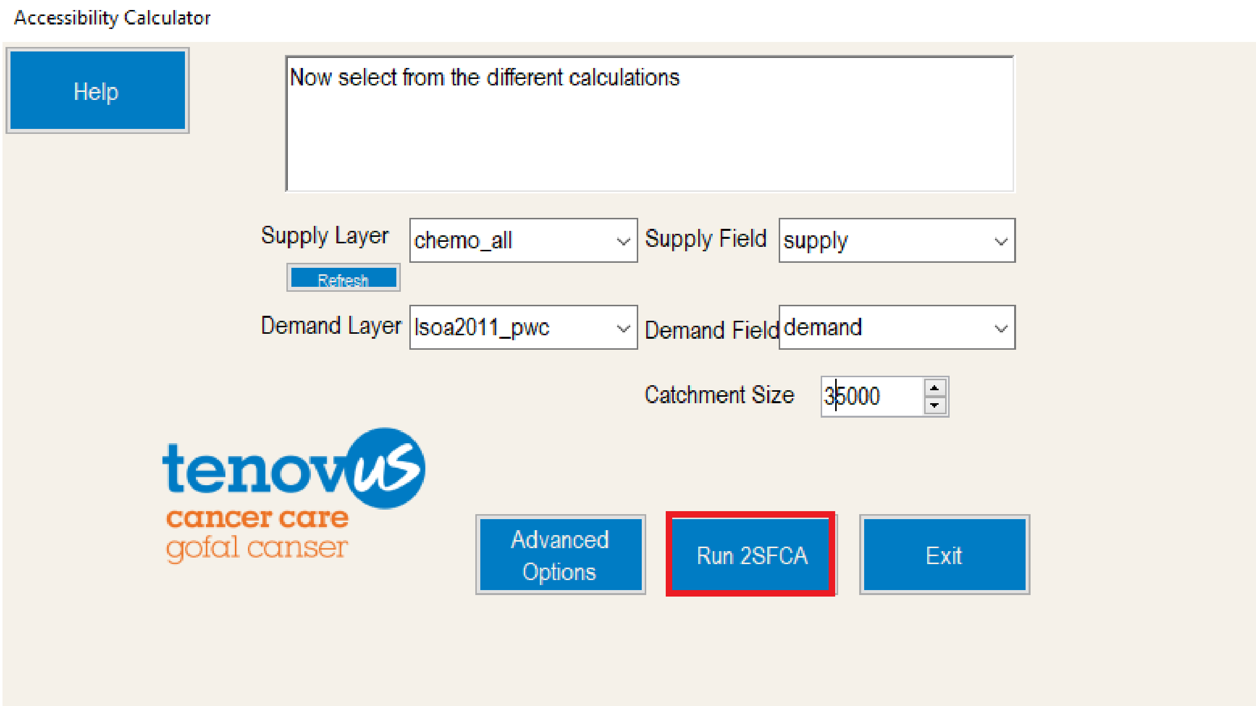


The demand field needs to be completed to show which column from the database table is required and, in this example, demand is selected. This can vary depending upon the data that is being used.

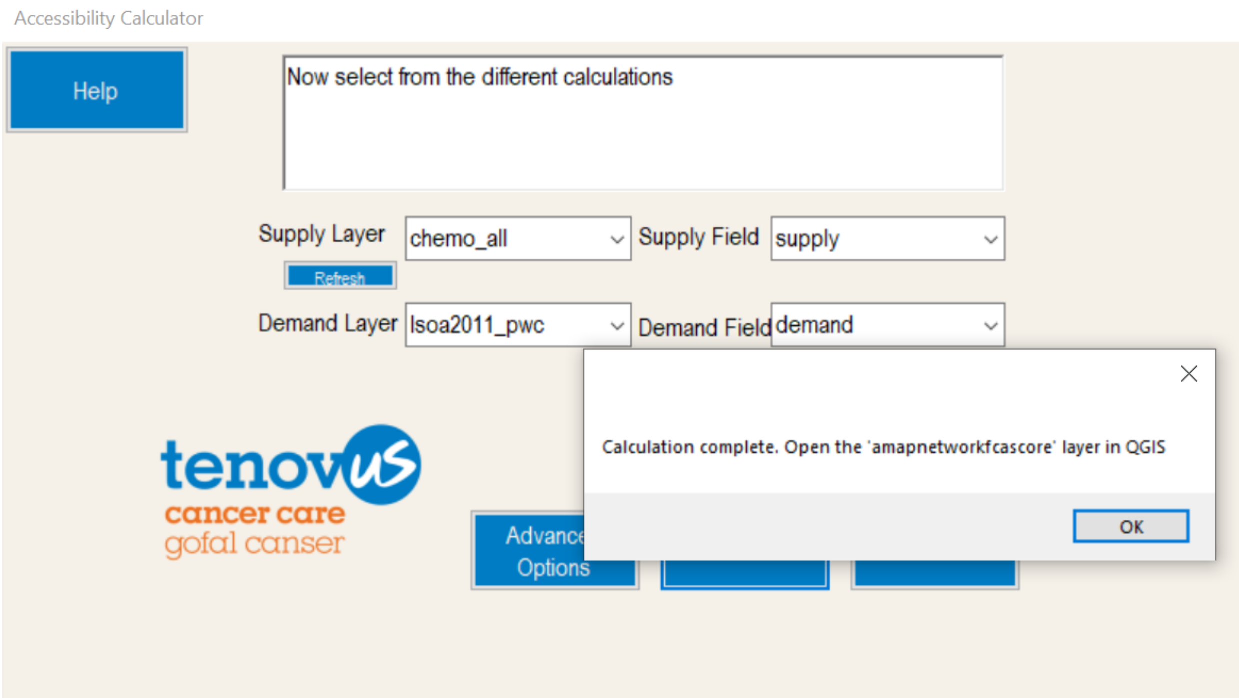
Accessibility Calculator



The last input is the distance which people travel; this can vary depending upon the services being supplied. It is likely that people will travel further for chemotherapy than they would to go to a dentist or an optician. This study found that the majority of people travelled less than 35km to use the Tenovus mobile units which is the distance being used in this example.

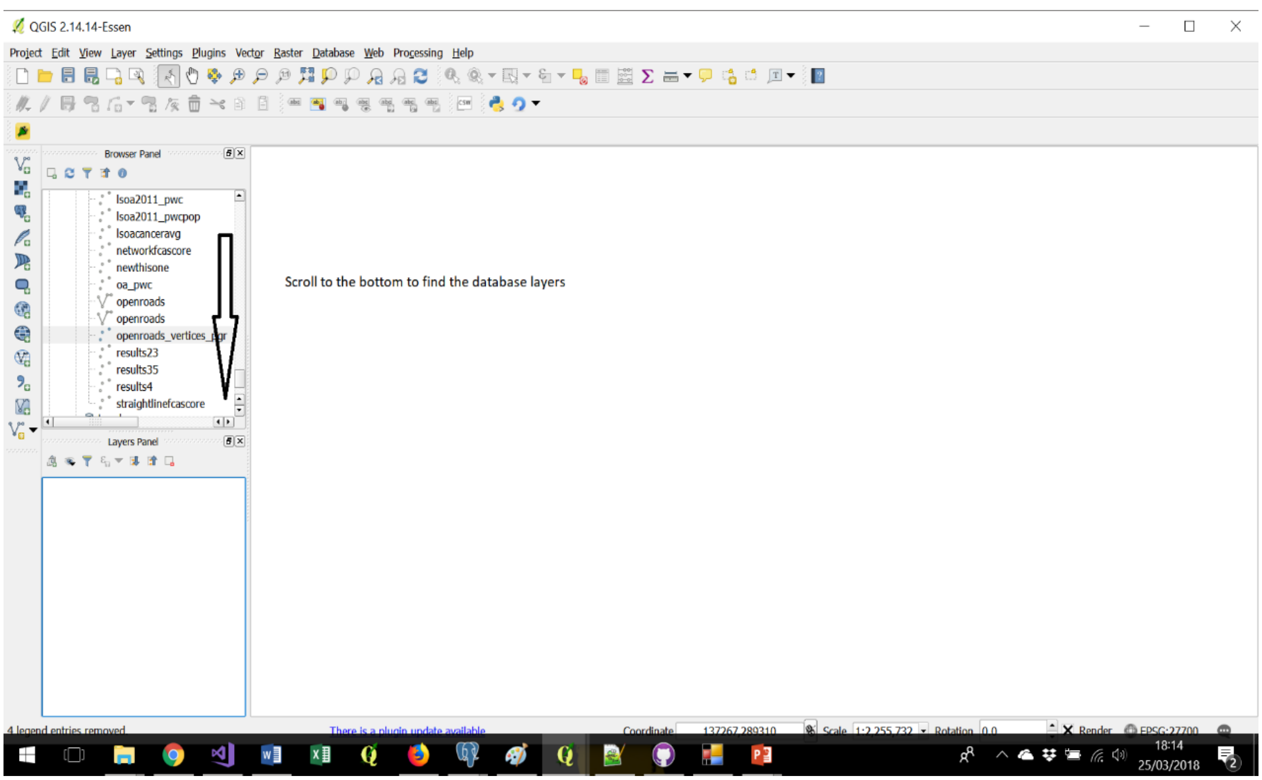


Pressing the Run 2SFCA button will start the program which will use a linear distance decay function (see section 3) and a networked distance as standard.

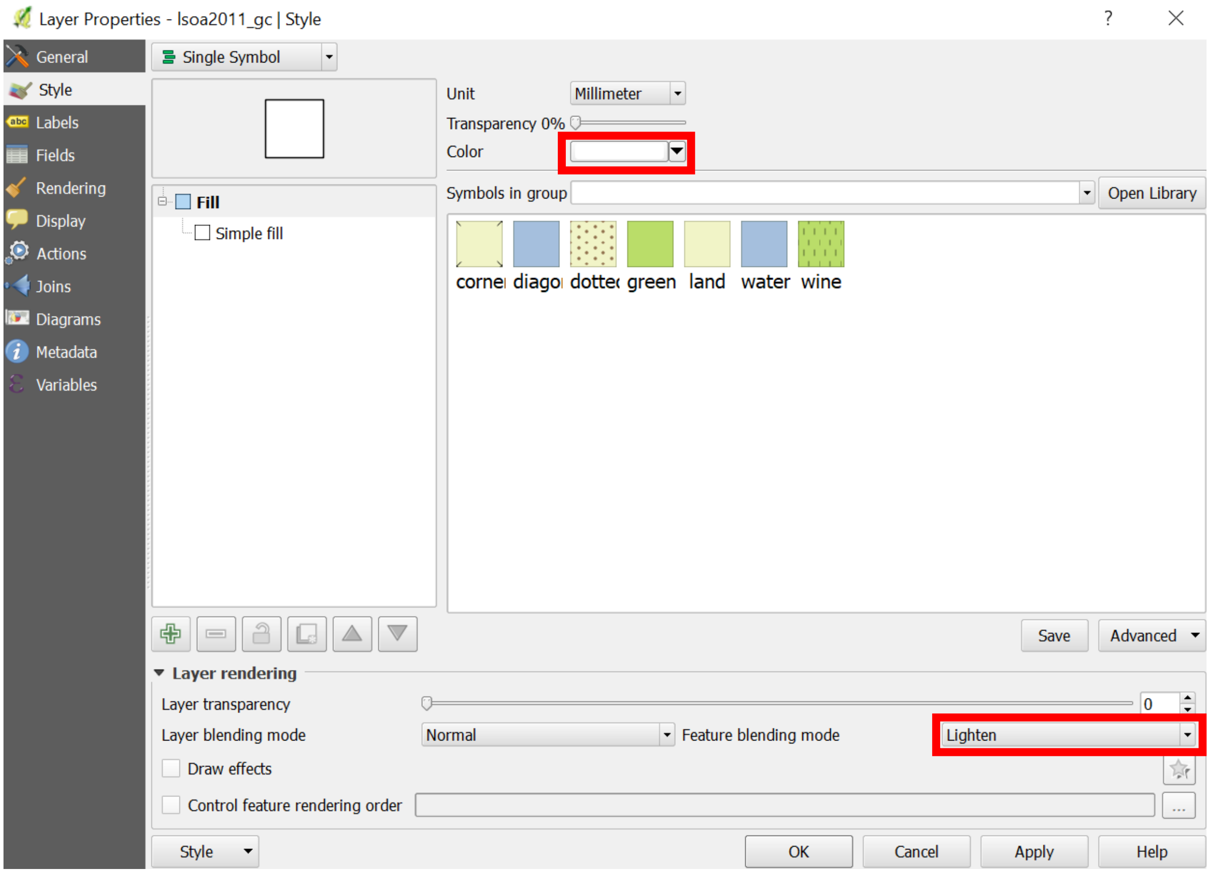


This pop-up shows that the tool has finished and the results are ready to visualised on a map.

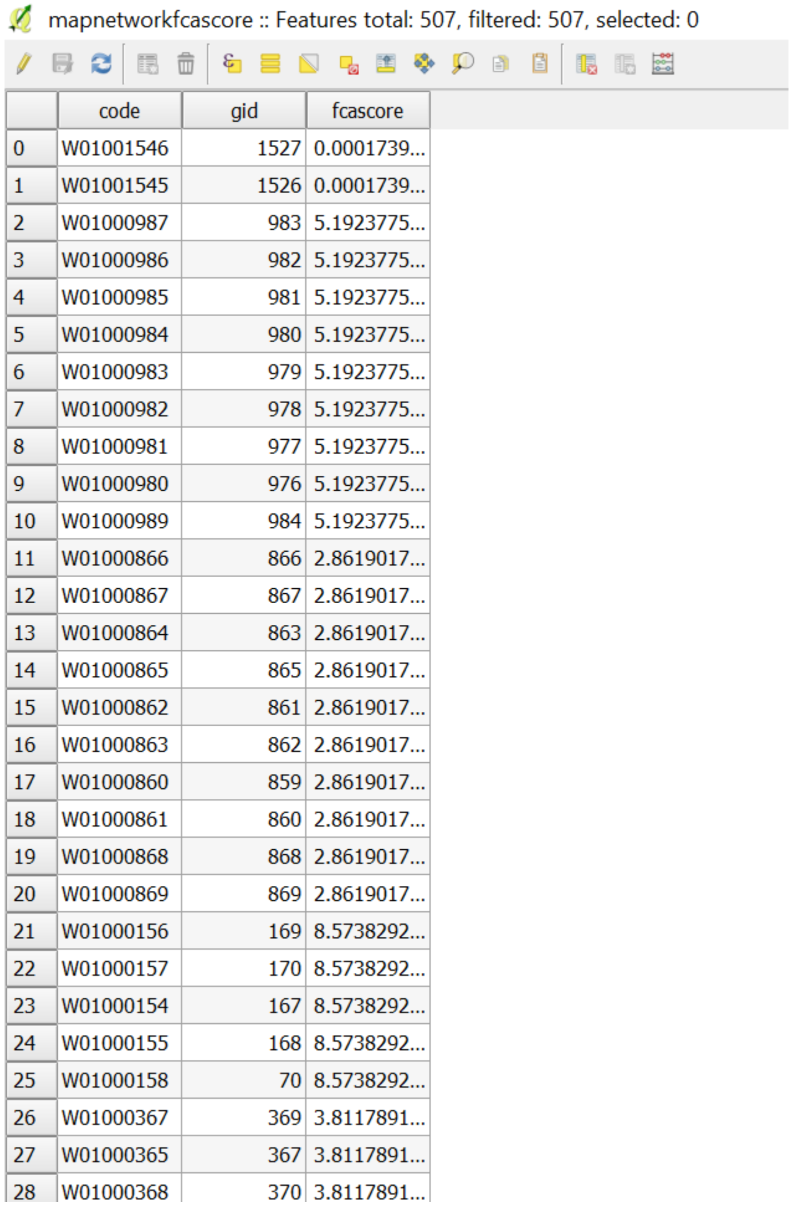
QGIS visualisation



Open the LSOA2011_gc layer – do this by double clicking on it, this layer has the boundaries of the LSOAs in Wales. Double click on the layer once it is open and change the colour to white. Change the feature blending mode to Lighten to remove the strong lines of each LSOA.



Open the amapnetworkfcascore table – do this by double clicking on it. This is the layer which holds all of the scores, if you right click on it and select attribute table it is possible to see the raw data we are going to display. The column of interest is fcascore which holds the 2SFCA scores

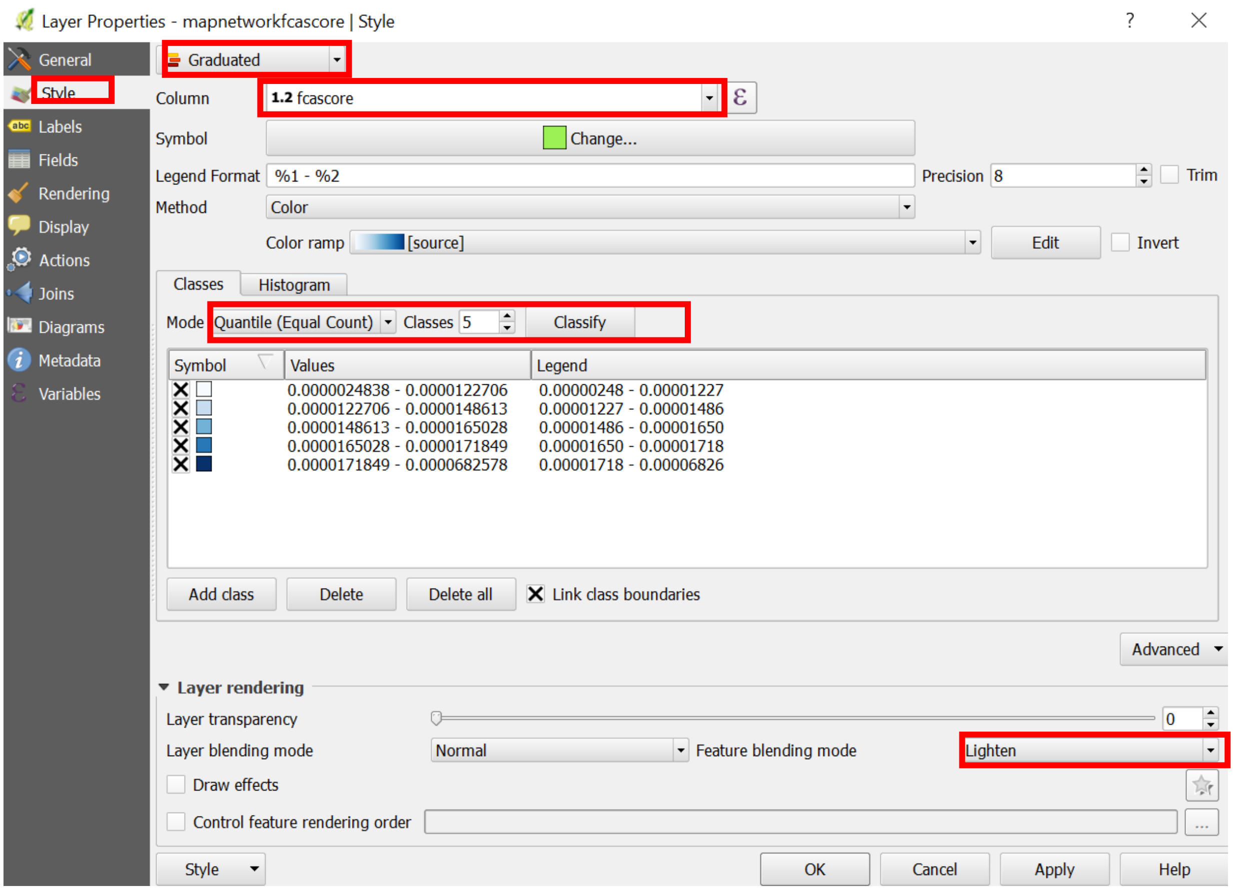


mapnetworkfcascore :: Features total: 507, filtered: 507, selected: 0

	code	gid	fcascore
0	W01001546	1527	0.0001739...
1	W01001545	1526	0.0001739...
2	W01000987	983	5.1923775...
3	W01000986	982	5.1923775...
4	W01000985	981	5.1923775...
5	W01000984	980	5.1923775...
6	W01000983	979	5.1923775...
7	W01000982	978	5.1923775...
8	W01000981	977	5.1923775...
9	W01000980	976	5.1923775...
10	W01000989	984	5.1923775...
11	W01000866	866	2.8619017...
12	W01000867	867	2.8619017...
13	W01000864	863	2.8619017...
14	W01000865	865	2.8619017...
15	W01000862	861	2.8619017...
16	W01000863	862	2.8619017...
17	W01000860	859	2.8619017...
18	W01000861	860	2.8619017...
19	W01000868	868	2.8619017...
20	W01000869	869	2.8619017...
21	W01000156	169	8.5738292...
22	W01000157	170	8.5738292...
23	W01000154	167	8.5738292...
24	W01000155	168	8.5738292...
25	W01000158	70	8.5738292...
26	W01000367	369	3.8117891...
27	W01000365	367	3.8117891...
28	W01000368	370	3.8117891...

Open the chemo_all layer – do the same as you did before, locate it and double click it to add it to the screen.

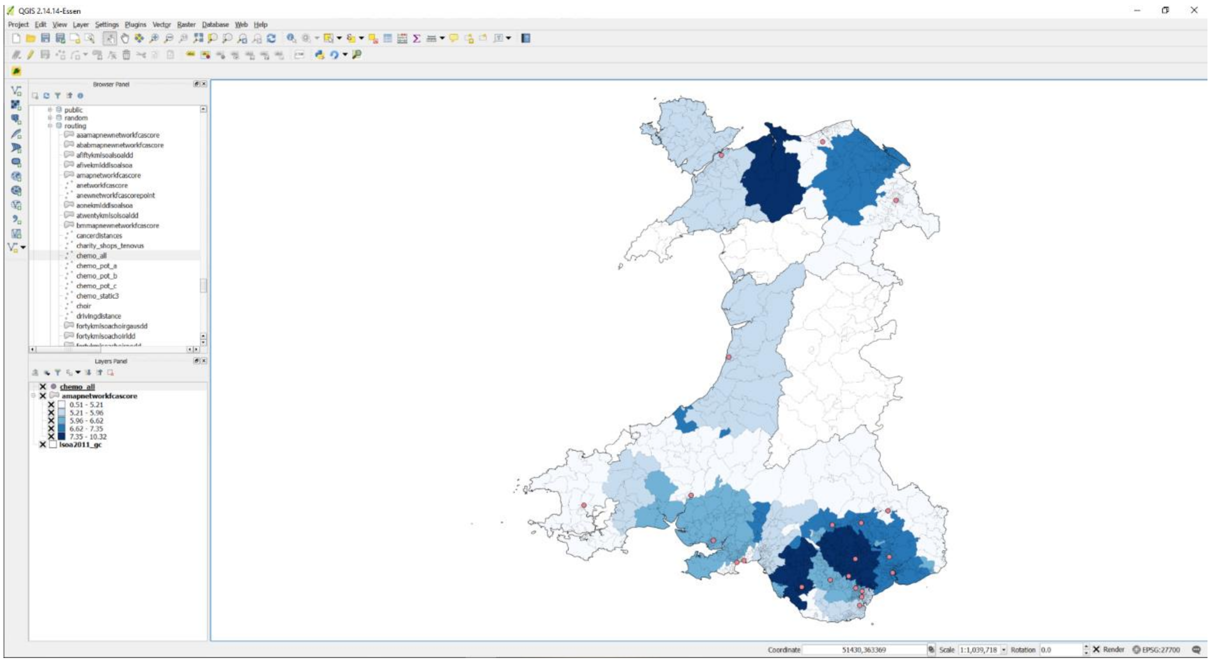
Map the data – All the data that is needed is in one location now, we finally need to visualise it in map form. Double click on the amapnetworkfcascore layer, and click on the style tab.



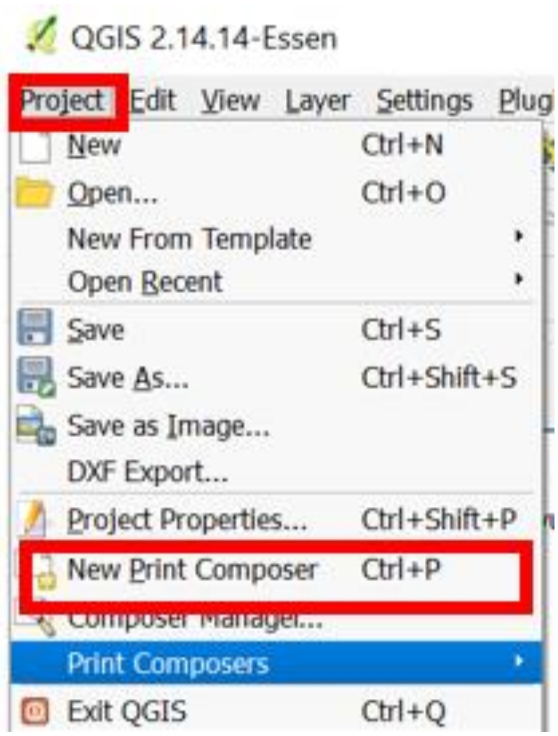
Graduate –there are many ways to visualise information using QGIS but for the best results with this type of data, it is better to follow the above. First, we select Graduated at the top, then we select the column that we want to visualise in this case it is the fcascore column. I like to use the blues colour ramp but there are many to experiment with, if you click on this option you can try and see which ones you think work best with the data. We then need to select the Quantile mode but again these can be experimented with and then press classify. The last thing we need to do on this page is to lighten the feature blending mode as it makes it easier to see the map. Once this is completed press the apply button and then OK.

Show original supply locations – drag the chemo_all layer to the top and it should be possible to see the effect that each choir is having on its surrounding.

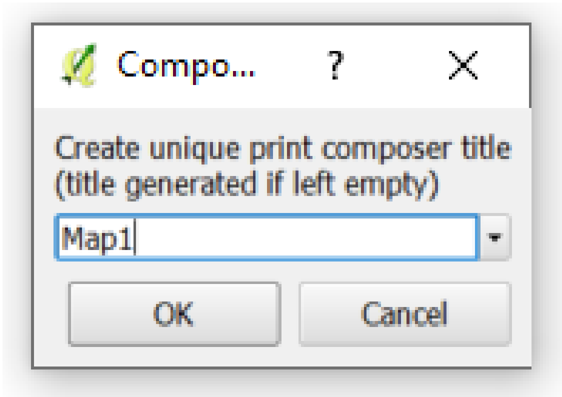
The map below shows what you should have, the darker the blue the better the accessibility. It is possible to see from the results that although the south east of Wales has many of the choirs there are still gaps where people have poor access to them. This may be as they are too far away or because they are sharing the choir with many other users.

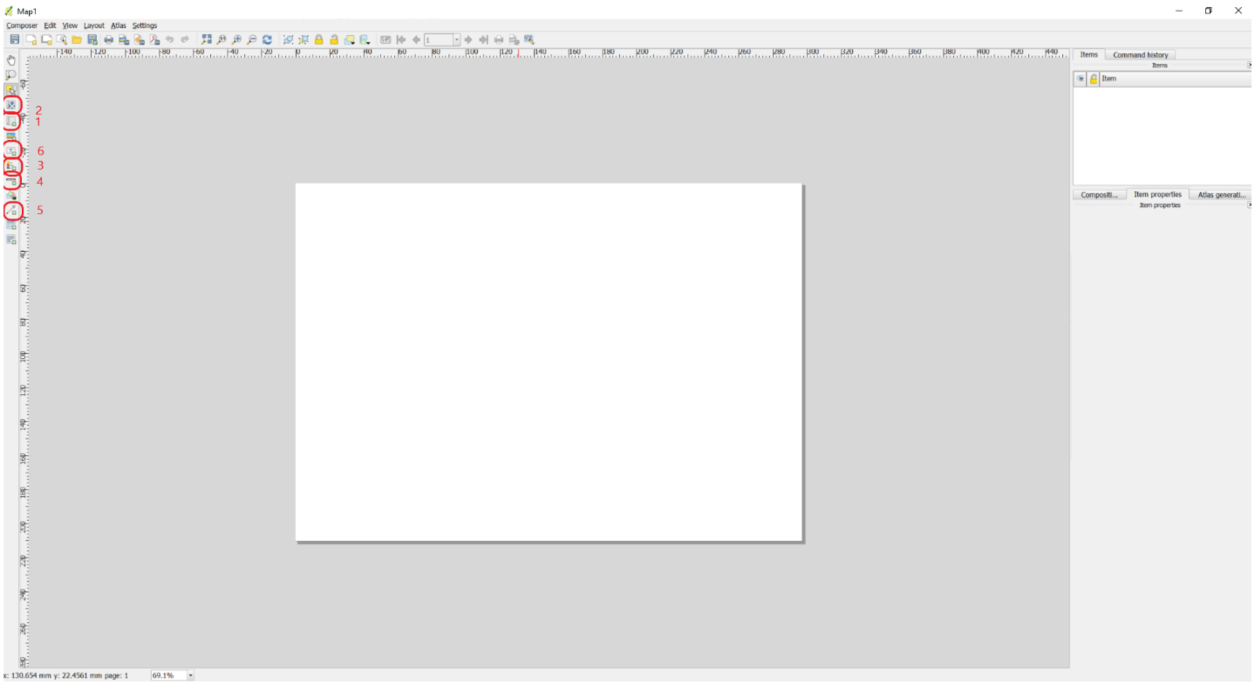


This is a good visual aid to understand the data but it is not well suited to sharing with others, there are several steps to creating a map and then saving it as a PDF or PNG. There are many good videos available on www.youtube.com which will show you some of the possibilities and below is a short guide to creating a map for presentation.



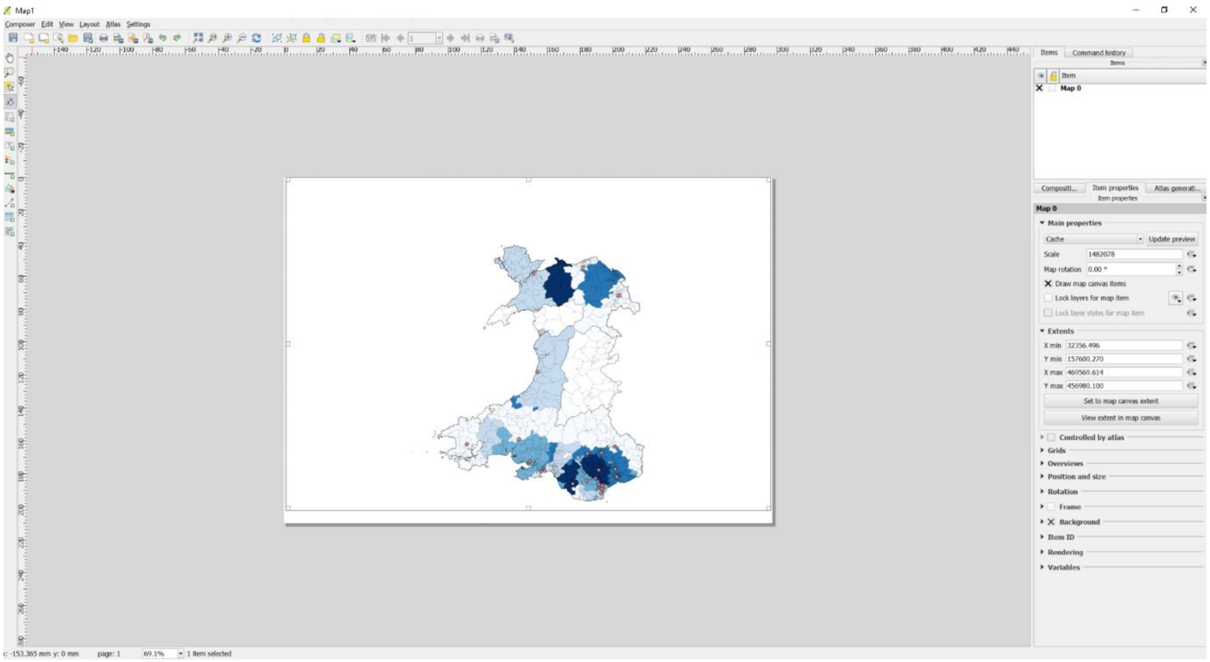
A name will need to be given to the composer and this can be called Map1.



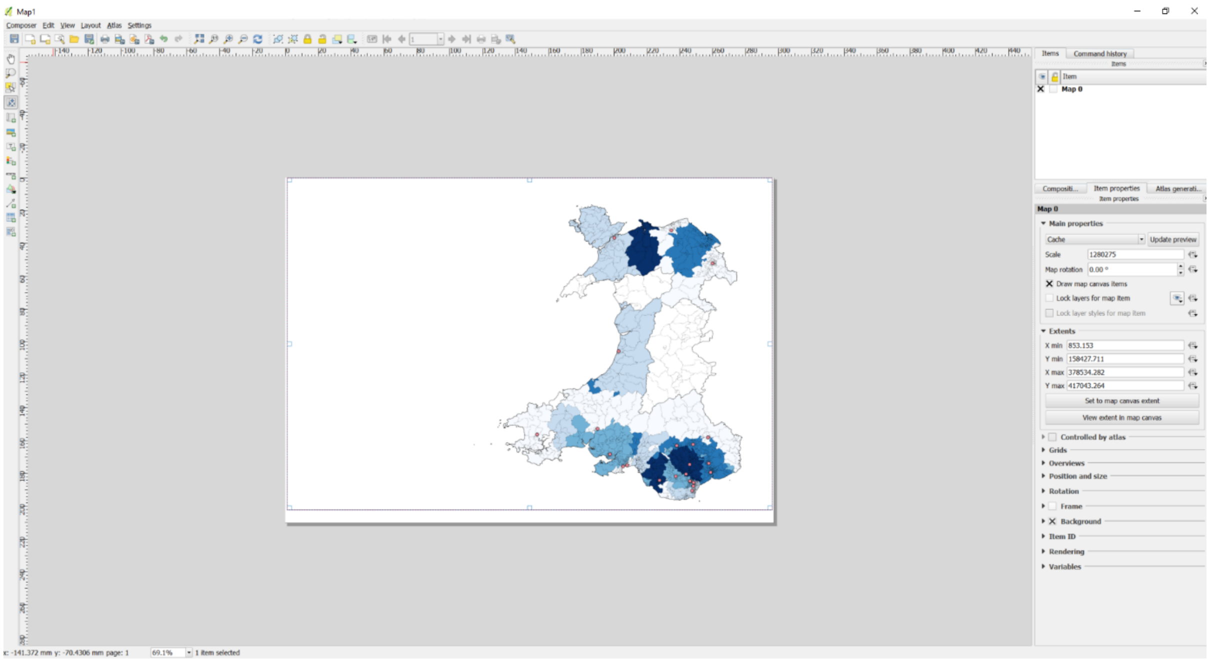


The above is the blank print composer screen that will be displayed. By using the buttons highlighted in red the map will be displayed.

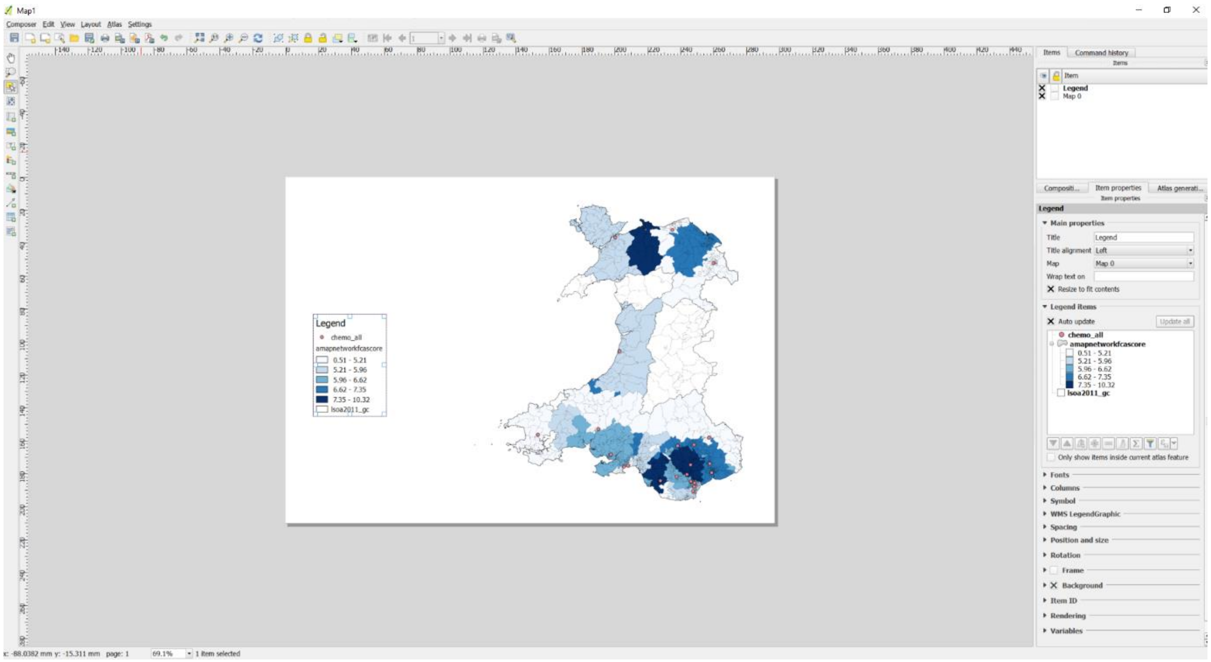
1. The first button to press brings the map from QGIS to the print composer. Once it is pressed the user needs to drag an area across the map and the map will be imposed upon it
2. This allows the user to centre the map in the frame
3. It is important to show the results and the legend will display the results of the graduation in QGIS
4. Every map should have a scale bar
5. Every map should have a North arrow
6. Every map should have a title which explains what it is displaying



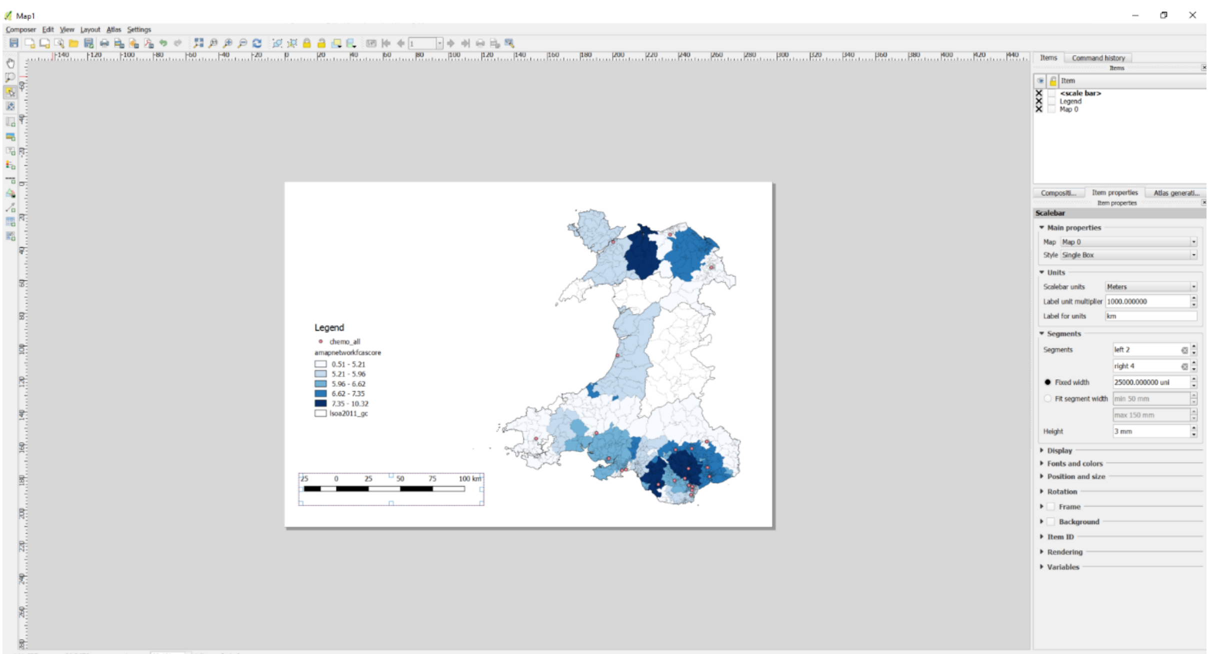
Drag a box across the map and then let go to show the map from QGIS.



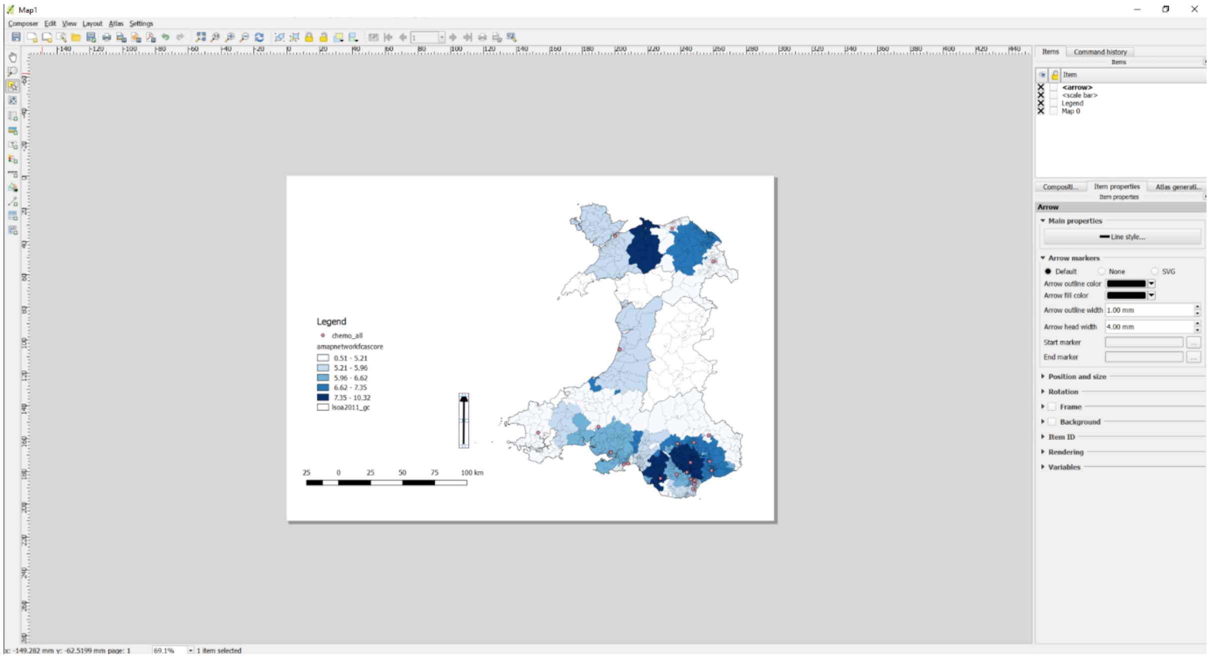
Using button 2 you can move the map on the canvas and make it bigger and smaller. Use the control key on the keyboard with the mouse wheel to make smaller size changes.



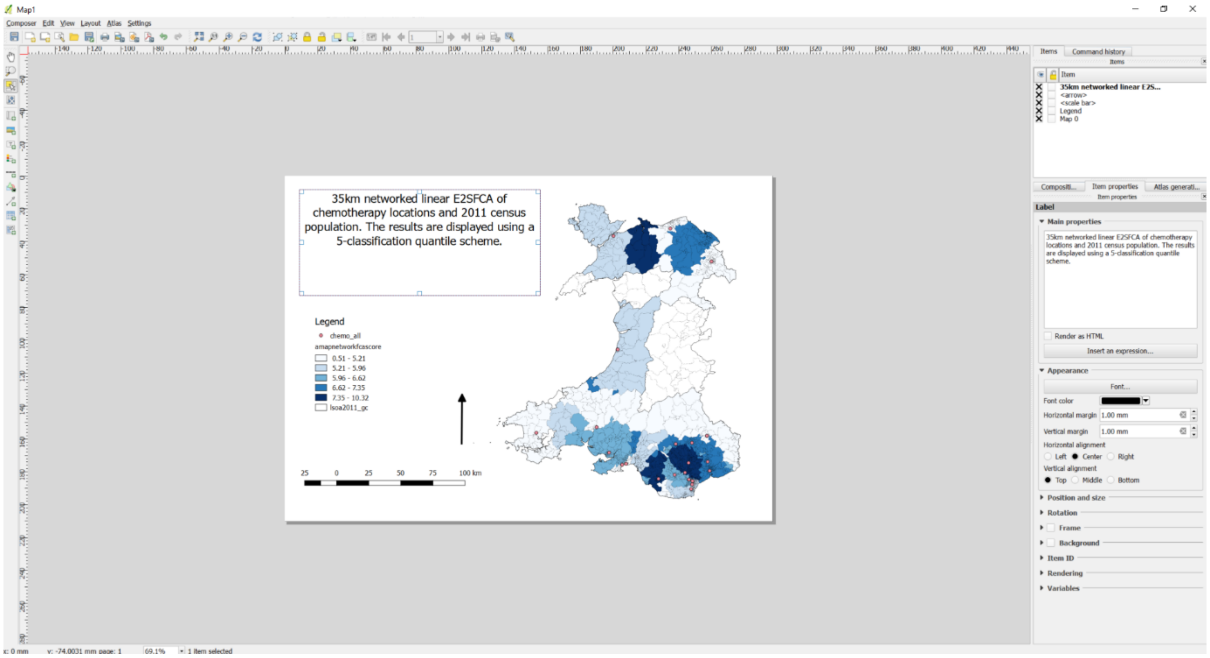
Insert the legend, it is good practice to change the names and tidy it up which can be done using the panel on the right of the screen.



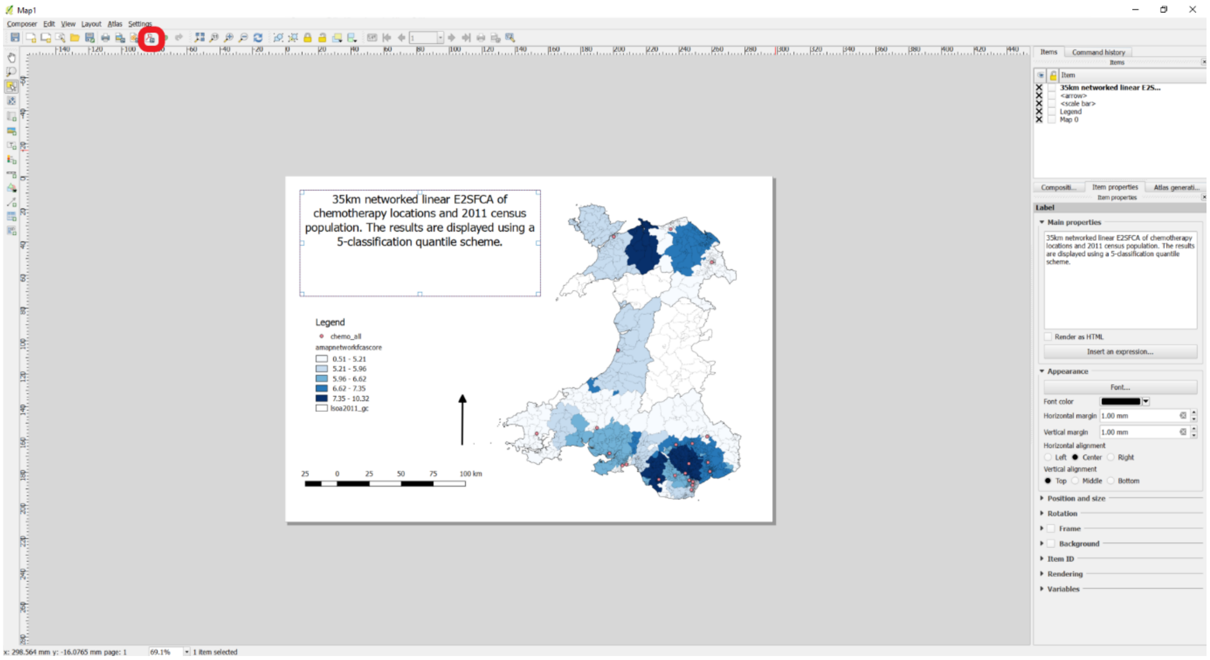
A scale bar needs to be inserted, there are multiple style choices that can be made and it can be made to show larger or smaller distances using the panel on the right.



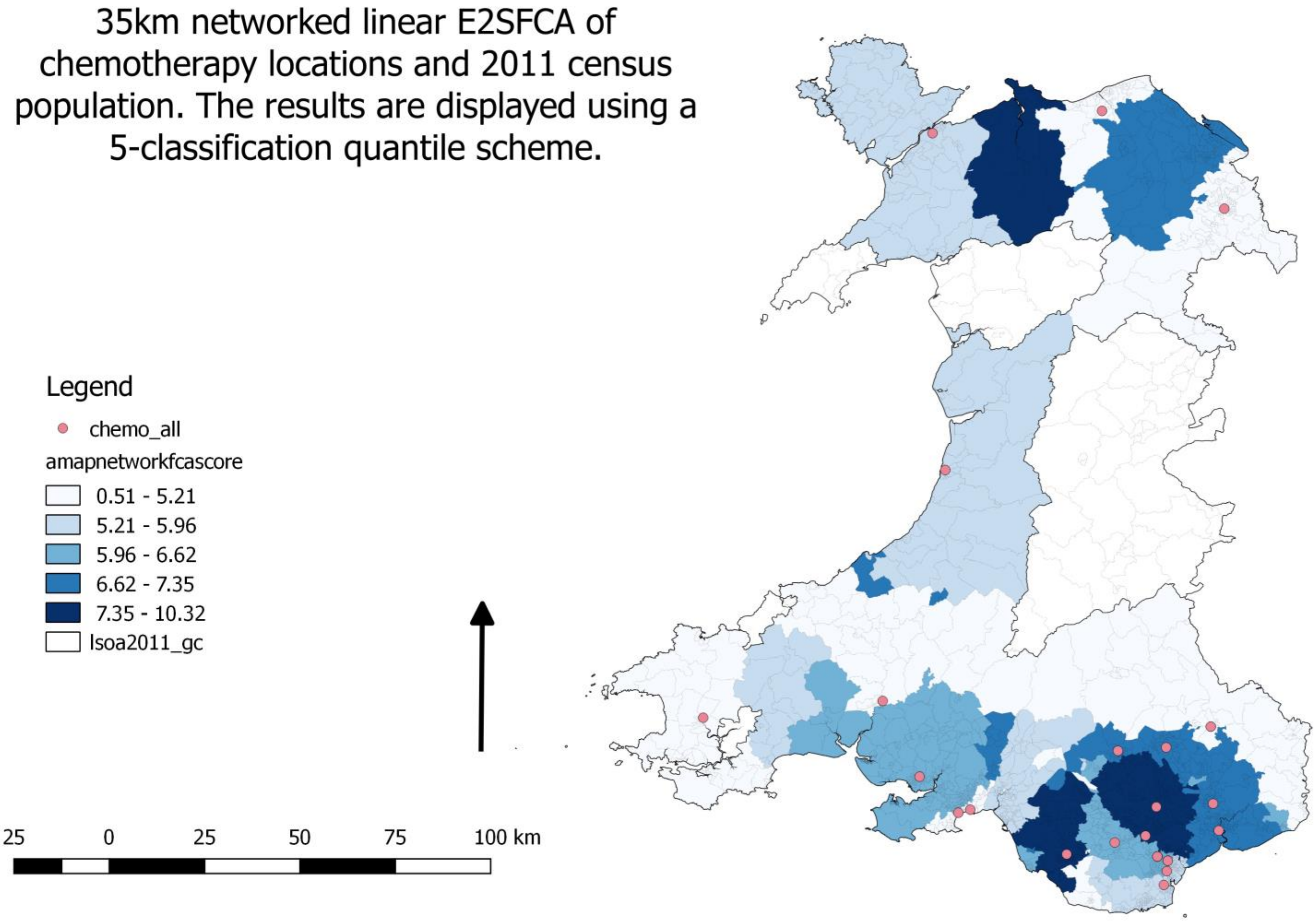
Insert the north bar into the image, this can be put wherever suits best. This can be customised and adjusted by using the panel on the right.



It is important to give each map a title which describes what is happening in the image. This one says the size of the catchment (35km), the type of analysis (linear E2SFCA), the supply data (chemotherapy locations) and the demand data (2011 census population). The font button on the right panel can be opened to adjust the font and size of the text.



The image can be saved as a PDF using the button highlighted in the top tool bar. Save the image in a sensible way which can help you to know what data is being displayed.



The final image is above and can be easily shared with others via printing or email.

C.2 Super user documentation

This document will outline the software required to run the tool, the process for installing them and how to download and transform the required data. The tool may work with other data and in terms of visualisation most GIS packages should work but for this document the tried and tested options will be explained.

Downloading & Installing software

The software required can be downloaded from OSGEO using https://download.osgeo.org/ there are other available sources. I would highly recommend using the PostgreSQL stack builder to do this as you can get most of the packages needed. The program should work with most of the versions of the software but the recommended versions are;

QGIS – v. 2.14 (for visualisation only)

PostgreSQL – v. 9.5

PostGIS – v. 2.3 (stack builder)

Microsoft Visual Studio (to change the code of the program)

pgRouting – v. 2.3.2 (stack builder)

NPGSQL (stack builder)

osm2po – v. 5.2.43 (for transforming OSM data and can be downloaded from: https://osm2po.de/)

Java runtime environment – (can be downloaded from https://www.java.com/en/download/)

2SFCA tool from download or GitHub

There is a good tutorial on the installation of the items 2,3,5, and 6 here: https://www.bostongis.com/PrinterFriendly.aspx?content_name=postgis_tut01

Updating the program

To do this it is best to use the visual studio IDE available from https://visualstudio.microsoft.com/downloads/. The code can be changed and it may be useful to change the pre-completed fields to the ones most suitable to your organisation. It is also possible to change the different methods and add functionality if required. There is a lot of documentation surrounding visual studio and it would be good to look at these when making updates.

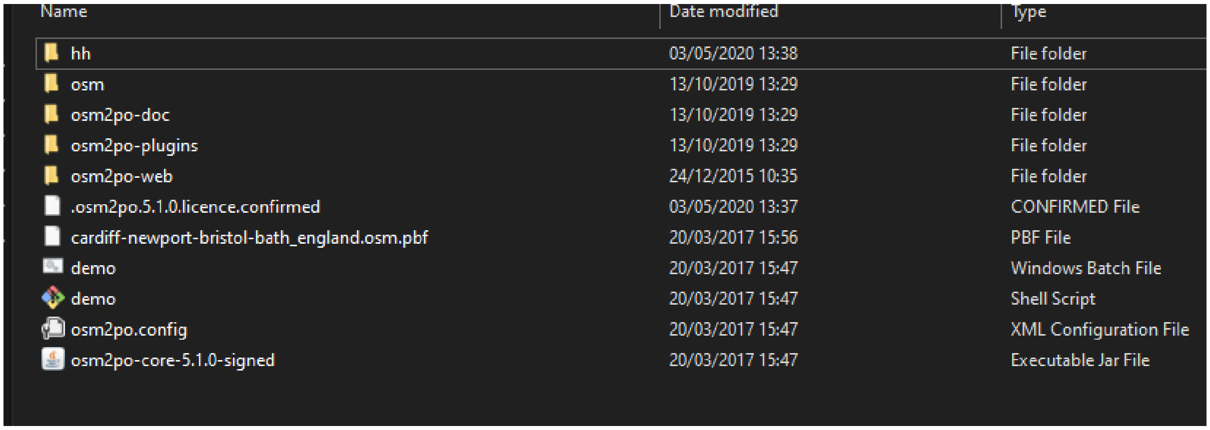
Downloading and transforming data

Network Data

Open Street Map Data

This can be quite a complex process depending on how experienced you are with computers. Java runtime will need to be installed before this process is started. The dataset for Wales can be downloaded from http://download.geofabrik.de/europe/great-britain.html. The .pbf format needs to be used in for the process to work. There are two good tutorials for this at https://www.bostongis.com/PrinterFriendly.aspx?content_name=pgrouting_osm2po_1 and another is https://anitagraser.com/2011/12/18/osm2po-part-2-pgrouting-on-osm-the-easy-way/. I would encourage you to read these first.

Initially copy the .pbf data file and move it to the osm2po folder.



Once the data has been downloaded and Java and osm2po are on your machine the command window needs to be opened. In the command window you will need to use a command like the one below;

D:\Downloads\osm2po-5.2.43>java -jar osm2po-core-5.2.43-signed.jar prefix=at "D:\Downloads\osm2po-5.2.43\wales.osm.pbf"

This will run osm2po and it may take a few minutes to complete depending on the dataset.

Once it is complete it is possible to see the data in your browser using http://localhost:8888/Osm2poService.

Next the data needs to be input in Postgres. A template can be found in the log file create.

```
hh_2po - Notepad
File  Edit  Format  View  Help
       Ways:32,236 Nodes:208,576
INFO   0 of 0 nodes cached (SHARED) - 901M
DEBUG  Hash Collisions Max : 0 Avg : 0.00
INFO   166,855 of 208,576 nodes cached (S090W180) - 874M
DEBUG  Hash Collisions Max : 6 Avg : 0.40
INFO   32,236 ways analyzed, 43,583 segments created (S090W180) - 836M
INFO   34,098 vertices of 166,855 nodes written - 836M
INFO   0 ways analyzed, 0 segments created (SHARED) - 836M
INFO   0 vertices of 0 nodes written - 836M
DEBUG  Writing TileFile:tm_info.2po
INFO   Segmenter finished at Sun May 03 13:37:40 BST 2020
TASK   Starting GraphBuilder at Sun May 03 13:37:40 BST 2020
INFO   Transforming Ways - 836M
INFO   32,236 Ways read, 87,168 Edges cached - 811M
INFO   Transforming Vertices - 811M
INFO   34,098 Vertices read, 3,010 Restrictions cached - 803M
INFO   Finding SourceVertex EntryPoints - 803M
INFO   87,168 Edges analyzed
INFO   Creating NoTurnBits - 797M
INFO   87,168 Edges analyzed, 5,694 Crossings checked
INFO   Writing GraphFile hh_2po.gph - 797M
DEBUG  GraphId is 1776942429
INFO   34,098 entries written
INFO   87,168 edges written
INFO   GraphBuilder finished at Sun May 03 13:37:41 BST 2020
TASK   Starting PostProccessor[0] at Sun May 03 13:37:41 BST 2020
       de.cm.osm2po.plugins.postp.PgRoutingWriter
INFO   Creating sql file hh\hh_2po_4pgr.sql
INFO   43,583 Segments written.
INFO   commandline template:
       psql -U [username] -d [dbname] -q -f "C:\Users\Rich\Desktop\Dropbox\PHD\osm2po\hh\hh_2po_4pgr.sql"
INFO   PostProcessor finished at Sun May 03 13:37:42 BST 2020
LOG    Closed at Sun May 03 13:37:42 BST 2020

LOG    Opened at Sun May 03 13:38:19 BST 2020
INFO   Loading library osm2po-demos-5.1.0.jar
INFO   Loading library osm2po-plugins-5.1.0.jar
INFO   Loading library osm2po-sd-5.1.0.jar
DEBUG  Registered Properties in
       jar:file:/C:/Users/Rich/Desktop/Dropbox/PHD/osm2po/osm2po-core-5.1.0-signed.jar!/osm2po.selfreg:
       service.mime.css=text/css
       service.mime.gif=image/gif
       service.mime.gz=application/gzip
       service.mime.html=text/html
       service.mime.jpg=image/jpg
       service.mime.js=text/javascript
```

The command looks like, you will need to update with your information;

psql -U postgres -d routing -q -f "D:\Downloads\osm2po-5.2.43\at\at_2po_4pgr.sql"

The data has been uploaded in the wrong format for the tool, the following command needs to run to change it from a multilinestring to a linestring. The section highlighted in red needs to be your schema.tablename.

create table public.hh_2po_4pgr1 AS

select id as gid, x1, y1, x2, y2, (ST_Dump(geom_way)).geom As geom from public.at_2po_4pgr;

The data has been imported in the wrong SRID for this to work, it will also be in the public schema and have some column names which need to be changed. To do use the below SQL command and it will make the amendments.

CREATE TABLE routing.osm_poa AS

SELECT gid, osm_id, osm_name, osm_source_id, osm_target_id, clazz, flags, source,

```
 target, kmh, km, cost AS cost_len, reverse_cost AS rcost_len,
 ST_X(ST_StartPoint(ST_Transform(geom, 27700))) as x1,
 ST_Y(ST_StartPoint(ST_Transform(geom, 27700))) as y1,
 ST_X(ST_EndPoint(ST_Transform(geom, 27700))) as x2,
 ST_Y(ST_EndPoint(ST_Transform(geom, 27700))) as y2,
 ST_Transform(geom, 27700) as geom
FROM public.at_2po_4pgr
LEFT JOIN public.hh_2po_4pgr1
ON public.hh_2po_4pgr1.gid = public.at_2po_4pgr.id ;
```

The part on the last line highlighted in red needs to be changed to your schema.tablename.

The final part is to add 2 additional columns;

```
ALTER TABLE routing.osm_poa

ADD COLUMN cost_time double precision,

ADD COLUMN rcost_time double precision,
```

The columns then need to be filled using the following;

```
UPDATE my_schema.hw_roadlink

SET cost_time = ST_Length(centrelinegeometry)/1000.0/speed_km::numeric*3600.0,

rcost_time = ST_Length(centrelinegeometry)/1000.0/speed_km::numeric*3600.0;
```

This should have created a routable network in the routing schema on PostGIS which can be accessed by the program.

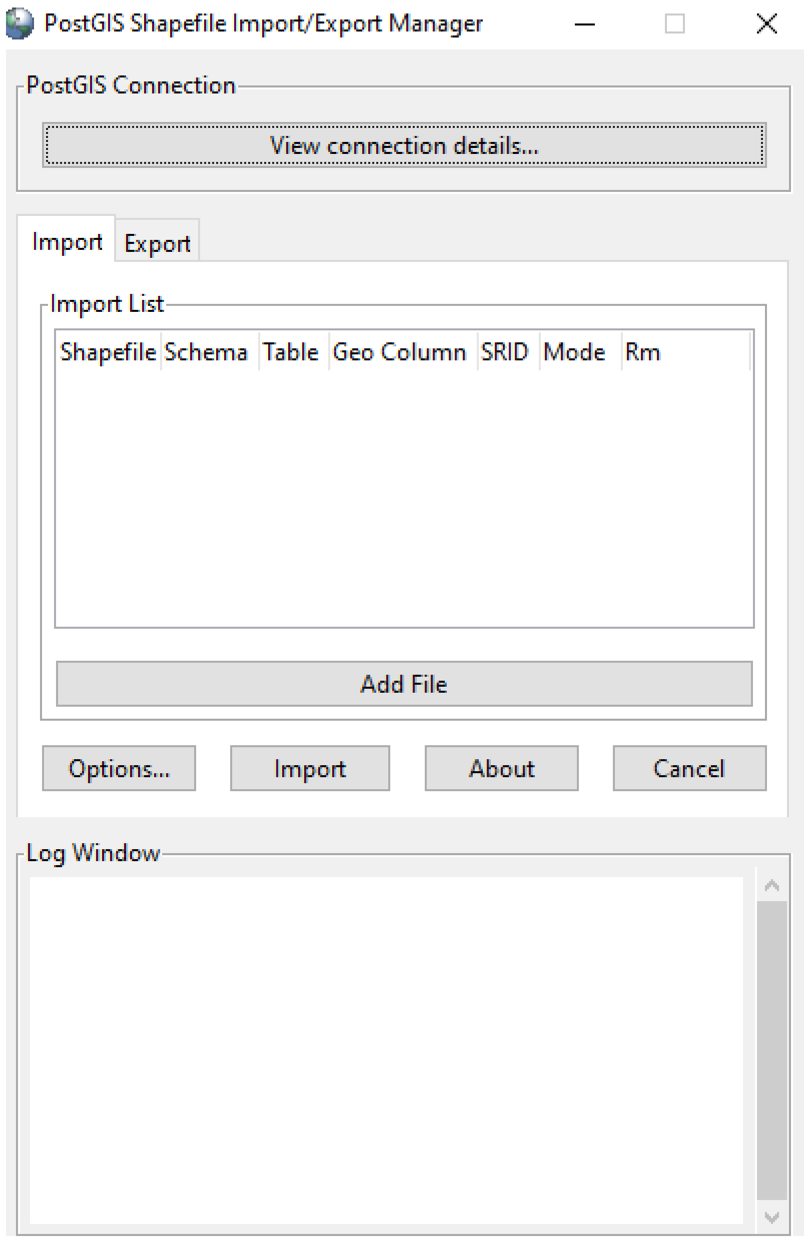
This is an example of the finished table;

	gid integer	osm_id bigint	osm_name character varying	osm_source_id bigint	osm_target_id bigint	clazz integer	flags integer	source integer	target integer	cost_len double precision	kmh integer	cost_time double precision	rcost_time double precision	x1 double precision	y1 double precision	x2 double precision	y2 double precision	geom geometry
1	120327	112183358	A55	1275927372	9711470	13	3	107895	107896	1.4461307	112	0.0129119	1000000	267976.378821247	375237.886006404	266998.695903049	374197.371100832	0102000020346C000011000000 8AB7E983215B10414D43458B1
2	120328	112183359	A55	1275927377	9689118	13	3	107897	107891	0.3794829	90	0.0042165	1000000	268098.085082905	375315.174327088	268444.826998683	375471.747109211	0102000020346C000004000000 1EF91F57085D1041DBCC82B24
3	120329	112183360	A55	9712709	1275927395	13	3	107893	107894	0.4736256	112	0.0042288	1000000	268458.666049909	375463.606185604	268027.252027885	375265.407475151	0102000020346C000005000000 C8FC08AAAA6210414AEBBB6C9
4	120330	112183362	A55	1275927399	1275927377	13	3	107898	107897	0.1581866	90	0.0017576	1000000	267957.279057616	375242.389389376	268098.085082905	375315.174327088	0102000020346C000004000000 A047C11DD55A1041A916BC8E2
5	120331	112183364	A55	9689123	1365473632	13	3	107892	107899	0.032181	90	0.0003576	1000000	268487.310279478	375491.256480188	268513.723365706	375508.215399033	0102000020346C000003000000 48E7B93D1D63104115BEA2060
6	120332	112260431	Heldre Lane	1276480970	1276480854	43	3	107900	107904	0.7429266	50	0.0148585	0.0148585	328697.80284269	310808.720810236	328042.139079944	310621.07733548	0102000020346C000014000000 DE641C36E70F144116141CE26
7	120333	112260431	Heldre Lane	1276480854	1276480738	43	3	107904	39789	0.686415	50	0.0137283	0.0137283	328042.139079944	310621.07733548	328073.881086965	309979.080625288	0102000020346C00000E000000 8F96A8EA80514413508314F7
8	120334	112260431	Heldre Lane	1276480738	5111937752	43	3	39789	107901	0.4916813	50	0.0098336	0.0098336	328073.881086965	309979.080625288	328315.888979707	309563.117026801	0102000020346C000012000000 4FA93B8627061441796F8F526
9	120335	112260434		1276480901	1072883477	43	3	107902	91871	0.1106168	50	0.0022123	0.0022123	328978.32289546	311954.922393545	329005.730894565	311847.702875579	0102000020346C000003000000 881BA54A49141441FDEE87B04
10	120336	112260437		11246244	1276480854	43	3	107903	107904	0.5807241	50	0.0116145	0.0116145	327573.237602913	310926.103240537	328042.139079944	310621.07733548	0102000020346C000010000000 982D4EF354FE134122E3B7693
11	120337	112260438	Heldre Lane	1276480942	1276480869	43	3	107905	107906	0.9391564	50	0.0187831	0.0187831	326911.161669243	309856.756773504	326090.789426523	309553.637030952	0102000020346C000013000000 379F8CA5FCF3134123A2EF068
12	120338	112260439	Parc Caradog	1276480809	1276480925	41	3	107907	107908	0.0630885	40	0.0015772	0.0015772	327267.434095728	310469.879796446	327233.589736505	310523.211982548	0102000020346C000002000000 2C9783BC8DF913410D5CE9841
13	120339	112260440	Burnt Lane	1276480745	2994435622	43	3	107909	107910	0.1576146	50	0.0031523	0.0031523	326922.41561629	310502.868753305	327052.590390918	310498.477076661	0102000020346C00000C000000 125197A929F4134160779A799
14	120340	112260441	Hazel Close	1276480759	1276480905	41	3	107911	107912	0.0478432	40	0.0011961	0.0011961	327346.262530047	310612.158158022	327310.794368725	310580.846112412	0102000020346C000005000000 35ADD40CC9FA1341292DF4A15
15	120341	112260447	Pecknall Lane	262892170	1276480751	43	3	107913	107914	3.5878783	50	0.0717576	0.0717576	334610.12409821	311671.318077133	334945.835341321	314641.658366007	0102000020346C000027000000 E899137F486C14411503B645D
16	120342	112260449	Parc Caradog	1276480906	4995658154	41	3	107915	155317	0.0295366	40	0.0007384	0.0007384	327175.608679908	310477.596259454	327202.385416261	310489.133286262	0102000020346C000003000000 25C9496F1EF8134195D691623
17	120343	112260449	Parc Caradog	4995658154	1276480925	41	3	155317	107908	0.0508782	40	0.001272	0.001272	327202.385416261	310489.133286262	327233.589736505	310523.211982548	0102000020346C000007000000 6C8FAA8A89F81341A3317C886
18	120344	112260449	Parc Caradog	1276480925	1276480800	41	3	107908	107923	0.0593651	40	0.0014841	0.0014841	327233.589736505	310523.211982548	327284.119193133	310552.996439695	0102000020346C000007000000 EDE2E35B06F9134104F411D9E
19	120345	112260449	Parc Caradog	1276480800	1276480905	41	3	107923	107912	0.0417337	40	0.0010433	0.0010433	327284.119193133	310552.996439695	327310.794368725	310580.846112412	0102000020346C000006000000 BCC30D7AD0F9134102B05AFC6
20	120346	112260449	Parc Caradog	1276480905	1276480840	41	3	107912	107916	0.041747	40	0.0010437	0.0010437	327310.794368725	310580.846112412	327291.233636151	310617.776184166	0102000020346C000003000000 BCFE6E2D3BFA1341CF4A6B62D
21	120347	112260450		60069417	2440803340	43	3	107917	107918	0.9511046	50	0.0190221	0.0190221	331421.451863264	313978.70375342	332115.066698165	314502.956442918	0102000020346C00000F000000 503EB5CE753A144189BCA4D0E
22	120348	112260452	Sale Lane	1276480908	1276480790	43	3	107919	107921	0.056548	50	0.001131	0.001131	327057.777872866	310500.858060624	327093.96392663	310458.170351625	0102000020346C000004000000 5AB48A1C47F61341B271A76E9
23	120349	112260452	Sale Lane	1276480790	1276480942	43	3	107921	107905	0.6598246	50	0.0131965	0.0131965	327093.96392663	310458.170351625	326911.161669243	309856.756773504	0102000020346C000006000000 20950FDBD7F6134108A870AEE
24	120351	112260458	Parc Caradog	1276480801	1276480800	41	3	107922	107923	0.0951582	40	0.002379	0.002379	327328.760425922	310472.044521708	327284.119193133	310552.996439695	0102000020346C000007000000 CA17AD0A83FA13414019972D2
25	120352	112260459	Heldre Lane	1276480738	1276480942	43	3	39789	107905	1.3357871	50	0.0267157	0.0267157	328073.881086965	309979.080625288	326911.161669243	309856.756773504	0102000020346C000009000000 4FA93B8627061441796F8F526
26	120353	112260794		2440810610	3431684739	43	3	107924	107925	0.0708616	50	0.0014172	0.0014172	332211.456372927	315351.40199282	332143.961446978	315372.861069735	0102000020346C000009000000 B86C53D3CD4614417A01A49B5
27	120354	112260795	Melverley View	1276486376	1276486444	41	3	107926	107927	0.1020452	40	0.0025511	0.0025511	333386.137356474	315234.226028543	333344.796177257	315296.652300998	0102000020346C000003000000 F02CA78C28591441C80674E78
28	120355	112260797	Belin Mount	1276486341	3431725749	41	3	107928	151935	0.0301207	40	0.000753	0.000753	332388.782023552	315595.996861392	332365.31708413	315590.427226257	0102000020346C000007000000 38C8CA2093491441903BC9FC2
29	120356	112260797	Belin Mount	3431725749	1276486379	41	3	151935	107929	0.0332798	40	0.000832	0.000832	332365.31708413	315590.427226257	332369.99708533	315557.463277959	0102000020346C000004000000 C4B3B14435491441CDCC7AB51
30	120357	112260800	Bryn Mawr	1276486456	392943259	41	3	107930	12422	0.0563522	40	0.0014088	0.0014088	332711.5289052	315496.51962708	332713.654974549	315545.131668763	0102000020346C000004000000 1C53991D9E4E1441111F1914A
31	120358	112260802		1276486421	1276486322	41	3	107931	107932	0.1011107	40	0.0025278	0.0025278	332295.248119953	315444.133281601	332256.611903217	315537.627801894	0102000020346C000003000000 332813FE1C481441D5F87A88D
32	120359	112260803	Maes Hafren	1276486394	1276486428	41	3	107933	107934	0.1375871	40	0.0034397	0.0034397	332081.455434661	315368.550859963	332113.427493132	315467.489564718	0102000020346C00001C000000 B5765DD2C54414415EA21434A
33	120360	112260804	Rodney Crescent	1276486385	1276486387	41	3	107935	107936	0.0052093	40	0.0001302	0.0001302	332203.745782534	315499.014431035	332208.8974974	315499.888147527	0102000020346C000002000000 A06AAEFBAE4614415602C70EA
34	120361	112260806	Malt Rise	1276486391	60069453	41	3	107937	107938	0.1523072	40	0.0038077	0.0038077	332269.205139499	315419.886614065	332145.286076969	315386.449300601	0102000020346C000019000000 BD1610D2B4471441BD8EE48B6
35	120362	112260807		1276486424	1276486442	41	3	107939	107940	0.129006	40	0.0032252	0.0032252	332173.228777207	315288.997948423	332089.363211986	315212.101602868	0102000020346C00000C000000 809244EA344614410131E6FD6
36	120363	112261253		1276492828	1276492852	43	3	107941	107942	0.0148123	50	0.0002962	0.0002962	299645.775502454	225898.226572713	299635.884890187	225909.284207792	0102000020346C000002000000 B2501D1AF7491241C65A05D05
37	120364	112261255		1276492806	1872514647	43	7	107943	132788	0.5258588	50	0.0105172	0.0105172	300183.889834828	224956.814819833	300035.538675154	225397.365796744	0102000020346C00001D000000 77DC308F5F521241B442C084E
38	120365	112261255		1872514647	1276492736	43	7	132788	132790	0.7479291	50	0.0149586	0.0149586	300035.538675154	225397.365796744	299660.814950171	225878.409510002	0102000020346C000020000000 9F759A270E501241DCD726EDA
39	120366	112261255		1276492736	1276492828	43	7	132790	107941	0.0250013	50	0.0005	0.0005	299660.814950171	225878.409510002	299645.775502454	225898.226572713	0102000020346C000003000000 2E4C8242334A1241042EAD46B
40	120367	112261257		1276492852	1028823443	43	3	107942	107944	0.1622251	50	0.0032445	0.0032445	299635.884890187	225909.284207792	299476.407375971	225937.633752009	0102000020346C000007000000 39A7208ACF4912412 2BC0E46A
41	120368	112261259		1028838937	378740353	43	3	107945	38134	0.182656	50	0.0036531	0.0036531	297760.933249234	226283.248774637	297629.454178698	226157.888077589	0102000020346C000007000000 F4AFA5BB832C12419E8E7DFD9
42	120369	112267611		1276572709	253856970	31	3	107946	13219	0.5243741	40	0.0131094	0.0131094	230382.932106044	377021.135782994	230346.239427411	377539.030781433	0102000020346C00000C000000 7A03F474771F0C4172B20A8BF
43	120370	112267611		253856970	253856980	31	3	13219	12066	0.621092	40	0.0155273	0.0155273	230346.239427411	377539.030781433	230436.41903684	378142.129212295	0102000020346C00000F000000 29EB58EA511E0C41F92A851F
44	120371	112267612	A4080	1276572659	1276572706	15	1	4635	107947	0.0256512	70	0.0003664	0.0003664	233541.647758838	372820.76453616	233526.548193007	372799.978288374	0102000020346C000002000000 672F9C2E2D820C412C91E20E5
45	120372	112267613		1276572673	1276572649	31	3	12065	107948	0.0305605	48	0.0006367	0.0006367	231044.604295538	376362.000101824	231020.917122302	376342.584168979	0102000020346C000002000000 02AE698D524340C414BB11A00A

Ordnance Survey OS Open Roads Data

The OS Open Roads data set is available for free as a download from https://www.ordnancesurvey.co.uk/business-government/products/open-map-roads this data set is easier to set up but less accurate and has less data than the OSM data. There are several good tutorials to look at like: https://ghost.mixedbredie.net/processing-thousands-of-journeys-a-pgrouting-function/.

First the data needs to be added to Postgres, to do this use the shapefile loader extension.



The SRID needs to be 27700 and the schema name will need to be updated in this example the schema is ‘routing’ and the table is ‘openroads’.

Once the data is in the required table the following SQL needs to be completed;

ALTER TABLE routing.openroads

ADD COLUMN source integer,

ADD COLUMN target integer,

ADD COLUMN speed_km integer,

```
 ADD COLUMN cost_len double precision,
 ADD COLUMN rcost_len double precision,
 ADD COLUMN cost_time double precision,
 ADD COLUMN rcost_time double precision,
 ADD COLUMN x1 double precision,
 ADD COLUMN y1 double precision,
 ADD COLUMN x2 double precision,
 ADD COLUMN y2 double precision,
 ADD COLUMN to_cost double precision,
 ADD COLUMN rule text,
 ADD COLUMN isolated integer;
```

The above command adds the required columns to the table, these columns will be filled in the following steps.

```
CREATE INDEX os_om_openroads_source_idx ON routing.openroads USING btree(source);
CREATE INDEX os_om_openroads_target_idx ON routing.openroads USING btree(target);
-- takes a few minutes


CREATE INDEX os_om_openroads_class_idx ON routing.openroads USING btree(class);
CREATE INDEX os_om_openroads_formofway_idx ON routing.openroads USING
btree(formofway);
-- takes a few more minutes
```

The above statements create the required indexes and increases the performance of the network.

```
UPDATE routing.openroads
 SET x1 = st_x(st_startpoint(geom)),
```

```
  y1 = st_y(st_startpoint(geom)),
  x2 = st_x(st_endpoint(geom)),
  y2 = st_y(st_endpoint(geom));


 UPDATE routing.openroads
  SET cost_len = ST_Length(geom),
  rcost_len = ST_Length(geom);
```

The above statements calculate the start and endpoint of each row and use the geom field to do this. It must be in 27700.

```
UPDATE routing.openroads SET speed_km =
 CASE WHEN class = 'A Road' AND formofway = 'Roundabout' THEN 20
 WHEN class = 'A Road' AND formofway = 'Collapsed Dual Carriageway' THEN 60
 WHEN class = 'A Road' AND formofway = 'Dual Carriageway' THEN 60
 WHEN class = 'A Road' AND formofway = 'Single Carriageway' THEN 55
 WHEN class = 'A Road' AND formofway = 'Slip Road' THEN 30
 WHEN class = 'B Road' AND formofway = 'Single Carriageway' THEN 50
 WHEN class = 'B Road' AND formofway = 'Collapsed Dual Carriageway' THEN 55
 WHEN class = 'B Road' AND formofway = 'Slip Road' THEN 30
 WHEN class = 'B Road' AND formofway = 'Roundabout' THEN 20
 WHEN class = 'B Road' AND formofway = 'Dual Carriageway' THEN 55
 WHEN class = 'Motorway' AND formofway = 'Collapsed Dual Carriageway' THEN 70
 WHEN class = 'Motorway' AND formofway = 'Dual Carriageway' THEN 70
 WHEN class = 'Motorway' AND formofway = 'Roundabout' THEN 20
```

```
 WHEN class = 'Motorway' AND formofway = 'Slip Road' THEN 30

 WHEN class = 'Motorway' AND formofway = 'Single Carriageway' THEN 60

 WHEN class = 'Not Classified' AND formofway = 'Roundabout' THEN 1

 WHEN class = 'Not Classified' AND formofway = 'Single Carriageway' THEN 30

 WHEN class = 'Not Classified' AND formofway = 'Slip Road' THEN 20

 WHEN class = 'Not Classified' AND formofway = 'Dual Carriageway' THEN 30

 WHEN class = 'Not Classified' AND formofway = 'Collapsed Dual Carriageway' THEN 30

 WHEN class = 'Unclassified' AND formofway = 'Single Carriageway' THEN 30

 WHEN class = 'Unclassified' AND formofway = 'Dual Carriageway' THEN 40

 WHEN class = 'Unclassified' AND formofway = 'Roundabout' THEN 20

 WHEN class = 'Unclassified' AND formofway = 'Slip Road' THEN 30

 WHEN class = 'Unclassified' AND formofway = 'Collapsed Dual Carriageway' THEN 40

 ELSE 1 END;
```

The above adds speed limits to the different classifications of roads within the dataset. These can be adjusted as is seen fit.

```
UPDATE routing.openroads

 SET cost_time = ST_Length(geom)/1000.0/speed_km::numeric*3600.0,

 rcost_time = ST_Length(geom)/1000.0/speed_km::numeric*3600.0;
```

Input a time cost and reverse cost.

```
ALTER TABLE routing.openroads

ALTER COLUMN geom TYPE geometry(MultiLineString)

USING ST_Force_2D(geom);
```

Ensuring the geometry type is correct.

```
SELECT pgr_createTopology('routing.openroads'::text, 0.001::float, 'geom'::text, 'gid'::text,
'source'::text, 'target'::text);
```

The above builds the topology which is what takes the data set and creates a routable network of nodes and lines.

```
SELECT pgr_analyzegraph('routing.openroads', 0.001, 'geom', 'gid', 'source', 'target');

VACUUM ANALYZE VERBOSE routing.openroads;
```

The above commands analyse the graph and look for errors.

```
SELECT * FROM pgr_drivingdistance('
    SELECT gid AS id,
        source::int4 AS source,
        target::int4 AS target,
        cost_len::float8 AS cost
    FROM routing.openroads',
    2000,
    10000,
    false,
    false)
```

2000 is the start node and 100000 is the distance

The last entry is a quick test which will output all the nodes within 10000 metres of node 2000. This will check that everything is working.

A vertices table will need to be created and the following command can do this;

SELECT pgr_createVerticesTable(routing.openroads, 'geom', 'source', 'target');

The documentation for this can be found here:
https://docs.pgrouting.org/2.0/en/src/common/doc/functions/create_vert_table.html

Supply Data

I have included 2 tables from Tenovus in the file which can be uploaded using the shapefile loader in PostGIS. These tables can be used a template for additional data to be loaded. There needs to be the following columns;

GID – an ID for each row

Geom – hold the location of each point, this may need to be renamed if it is geometry or the_geom

Supply – a value to be used as the supply, this could be named anything and there may be several columns with this data, number of beds, opening hours etc

Supply_nn – this is the nearest node to the location and will be filled with data in the program

Id– filled by program

Edge_id– filled by program

Pid– filled by program

Fraction – filled by program

An example of a supply table;

	gid [PK] serial	location f character varying(254)	unit character varying(254)	nurses_wte numeric	open_weekl numeric	beds_chair numeric(10,0)	supply numeric	capacity character varying(254)	postcode character varying(254)	x _east numeric(10,0)	y_north numeric(10,0)	geom geometry(Point,27700)	demand_nn integer	supply_nn integer	id integer	edge_id integer	pid integer	fraction double precision
1	1	Ysbyty Maelor Wre	no	9.1500000	42.500000	11	467.50	30	LL13 7TD	332375	350356	0101000020346C000		556000	1	65813	-1	0.8153399643
2	2	Ysbyty Glan Clwyd	no	9.8599999	42.500000	21	892.50	40	LL18 5UJ	300309	376019	0101000020346C000		603469	2	65770	-2	0.3062494249
3	3	Ysbyty Gwynedd	no	7.3100000	42.500000	20	850.00	30	LL57 2PW	255857	370173	0101000020346C000		643285	3	131395	-3	0.7637456239
4	4	Singleton Hospita	no	45.390000	50.000000	36	1800.0	55 (avg per day)	SA2 8QA	262626	191910	0101000020346C000		194792	4	131441	-4	0.7203371050
5	5	Royal Gwent Hospi	no	9.0000000	40.000000	26	1040.0		NP20 2UB	330932	187243	0101000020346C000		98546	5	85026	-5	0.3925228459
6	6	Nevill Hall Hospi	no	5.0000000	32.000000	11	352.00		NP7 7EG	328833	214506	0101000020346C000		323934	6	85149	-6	0.2376704775
7	7	Bronglais General	no	2.7600000	40.000000	6	240.00		SY23 1ER	259208	281805	0101000020346C000		361151	7	131145	11854'	1
8	8	Glangwili General	no	4.8000000	40.000000	8	320.00		SA31 2AF	242792	221240	0101000020346C000		368530	8	1400	-8	0.6377750944
9	9	Prince Philip Hos	no	3.2000000	40.000000	6	240.00		SA14 8QF	252459	201367	0101000020346C000		353725	9	99785	-9	0.6698480824
10	10	Withybush General	no	4.4000000	40.000000	8	320.00		SA61 2PZ	195710	216835	0101000020346C000		387417	10	109336	-10	0.5630533467
11	11	Chemo inpatient u	no	4.0000000	45.000000	14	630.00	6	CF14 2TL	314817	180435	0101000020346C000		2929	11	100026	-11	0.3371871764
12	12	Chemo day unit Ve	no	6.0000000	45.000000	12	540.00	30	CF14 2TL	314817	180435	0101000020346C000		2929	12	21140	-12	0.2467201037
13	13	Rhosyn day unit V	no	5.0000000	42.500000	9	382.50	25	CF14 2TL	314817	180435	0101000020346C000		2929	13	100017	-13	0.2026195221
14	14	Prince Charles Ho	no	5.0000000	40.000000	10	400.00	24	CF47 9DT	304481	208161	0101000020346C000		6320	14	107445	96066	0
15	15	Royal Glamorgan H	no	2.0000000	30.000000	3	90.000	12	CF72 8XR	303649	184110	0101000020346C000		128196	15	99851	-15	0.3762728494
16	16	Nevill Hall Hospi	no	5.0000000	40.000000	10	400.00	24	NP7 7EG	328833	214506	0101000020346C000		323934	16	103767	-16	0.6430797145
17	17	Princess of Wales	no	2.0000000	25.000000	3	75.000	12	CF31 1RQ	291047	181033	0101000020346C000		201995	17	99963	-17	0.8601129095
18	18	Ysbyty Ystrad Faw	no	2.0000000	15.000000	4	60.000	6	CF82 7EP	314564	193444	0101000020346C000		120066	18	27003	-18	0.7288222653
19	19	St Davids	no	2.0000000	27.500000	4	110.00	12	CF11 9XB	317297	176583	0101000020346C000		117664	19	99987	-19	0.1093358363
20	20	Ysbyty Aneurin	no	2.0000000	15.000000	3	45.000	6	NP23 6GL	317138	209045	0101000020346C000		327625	20	27035	-20	0.6682192705
21	21	University Hospit	no	24.000000	50.000000	27	1350.0		CF14 4XW	317551	179319	0101000020346C000		113058	21	23905	-21	0.5811762603
22	22	University Hospit	no	2.0000000	20.000000	4	80.000		cf64 2xx	316526	172981	0101000020346C000		118645	22	95479	-22	0.6986223427
23	23	Ward C4 Neuro Day	no	5.0000000	45.000000	7	315.00	12	CF14 4XW	317551	179319	0101000020346C000		113058	23	102727	-23	0.6719299576
24	24	Cwmbran	yes		8.0000000	7	56.000		NP44 3YS	329434	194301	0101000020346C000		101795	24	107631	-24	0.5677017138
25	25	Nantgarw	yes		8.0000000	7	56.000		CF15 7QX	311723	185868	0101000020346C000		121722	25	42920	-25	0.4023408924
26	26	Swansea	yes		8.0000000	7	56.000		SA1 3ST	265757	192701	0101000020346C000		194878	26	141973	-26	0.3822213430
*																		

Demand Data

I have included 2 tables of census data and this can be used as a template. The columns needed are the same as above. New data can be retrieved by using online resources such as https://census.ukdataservice.ac.uk/get-data.aspx. Any dataset that gives an insight into the people using the service will be good here and the data would be better at smaller sizes so LSOA is ideal.

2 3.2 SFCA Documentation

1 Introduction to healthcare accessibility

There are many ways for an organisation to decide where to put its resources. Within healthcare there are several commonly used methods including population to provider ratio (Figure 1), proximity (Figure 2) and a more complex solution called 2SFCA (figure 3). These are just a sub-section of the methods used by organisations but helps explain how the 2SFCA method works.

1.1 Population to provider ratio

One method used to measure accessibility is to count the amount of people within an area and then provide the services those people need. Figure 1 shows the different Welsh health boards; each one could calculate how many people live in its boundaries and then provide the amount of hospital beds for those people.

This works well but suggests that people do not move across these boundaries, and some of the health boards are quite large. For instance, if Powys decided to build a big single hospital in Newtown many of those who should use it may end up in either Cardiff or one of the north Wales hospitals.

Figure 1

1.2 Proximity

Proximity testing investigates how far people have to travel for a service and says that if people can access a service within a certain threshold then there is enough service. This can work well for some services but as demonstrated in Figure 2, just because someone is close to a service in a large area does not indicate that they have better access to that service than someone who has to travel further. For example, someone in Cardiff may have to travel a short distance to access their nearest hospital, but they are going to share that hospital with a large number of people. Whereas in Aberystwyth someone may need to travel further to reach the hospital but they then share it with less people and may get seen more quickly.

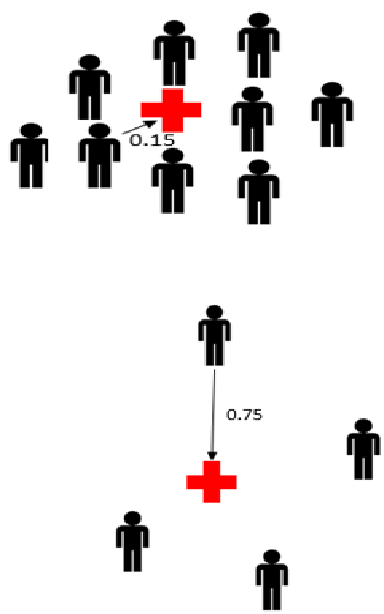


Figure 2

1.3 Benefits of 2SFCA

2SFCA tries to use both of these methods to understand not only the distance travelled to a service, but also the amount of people likely to access the service. It does this by completing the following steps:

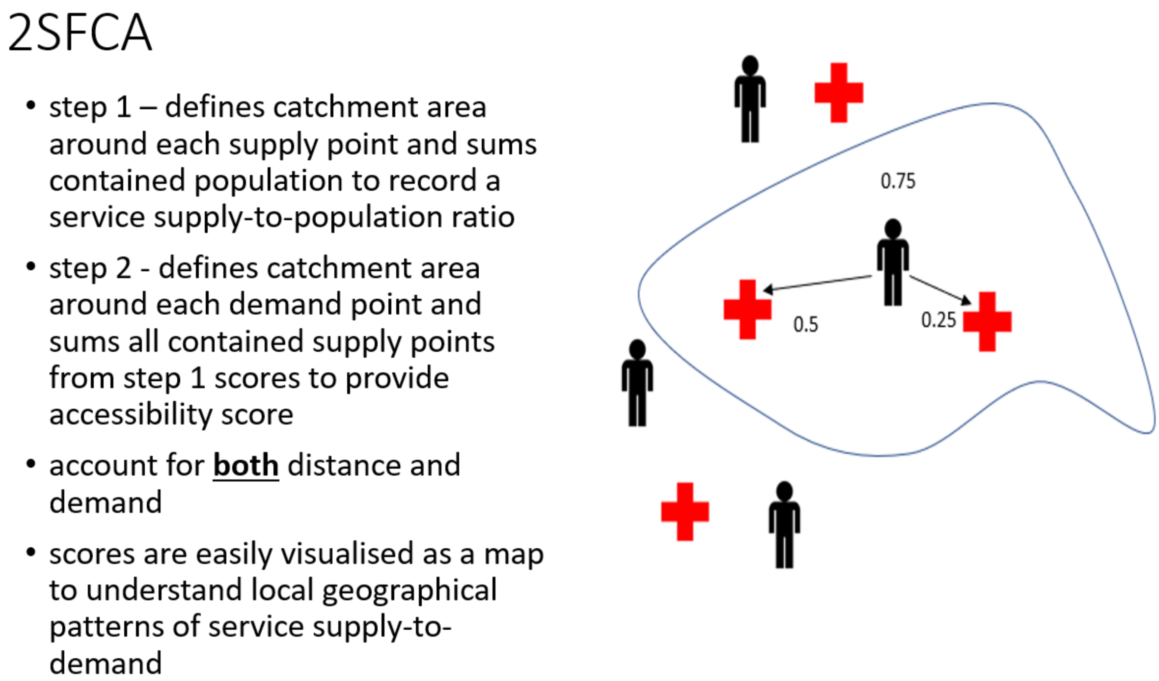


Figure 3

You do not need to understand how 2SFCA works in full but this should allow you an insight into why these types of studies are important. Many books delve much further into the subject and can assist in explaining how it works in detail.

1.4 Interpretations of 2SFCA results

Once the data has been visualised in QGIS a map should have been produced like figure 4 or 5. Figure 4 shows hospital locations in Wales with chemotherapy capacity. The results are banded into 6 classifications. The completely white area has no access to any chemotherapy unless they travel further than 35km. Then the colours are faded to show that the lighter the blue the poorer the access to the services. The dark blue areas on the map have the best accessibility in Wales.

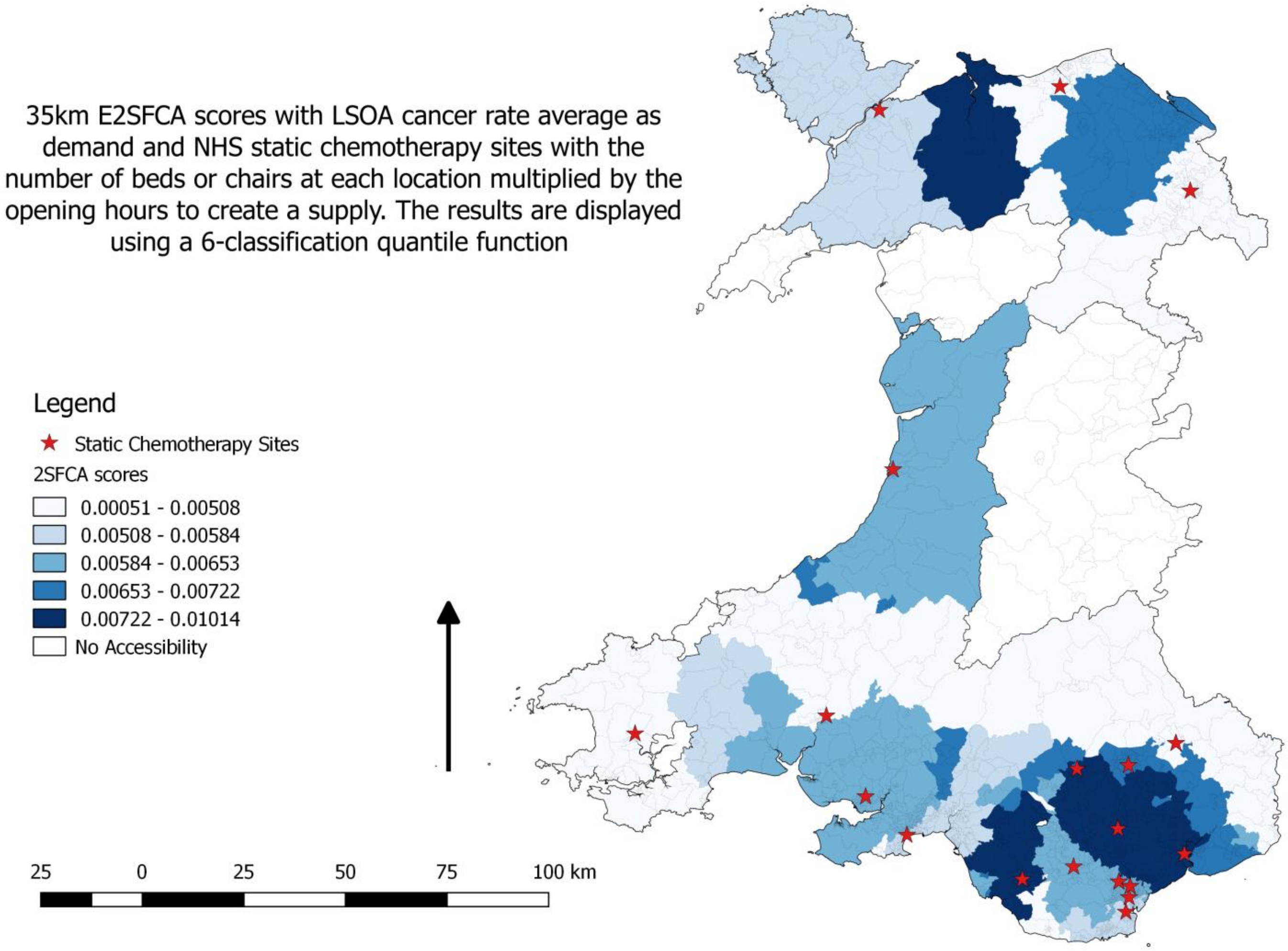


Figure 4

Figure 5 shows the accessibility people have to the Tenovus charity shops within Wales. Again, the colours are gradual and show that the dark blue areas have better access. So, west Wales has better access even though it has less shops, this is due to the size of the population that accesses them.

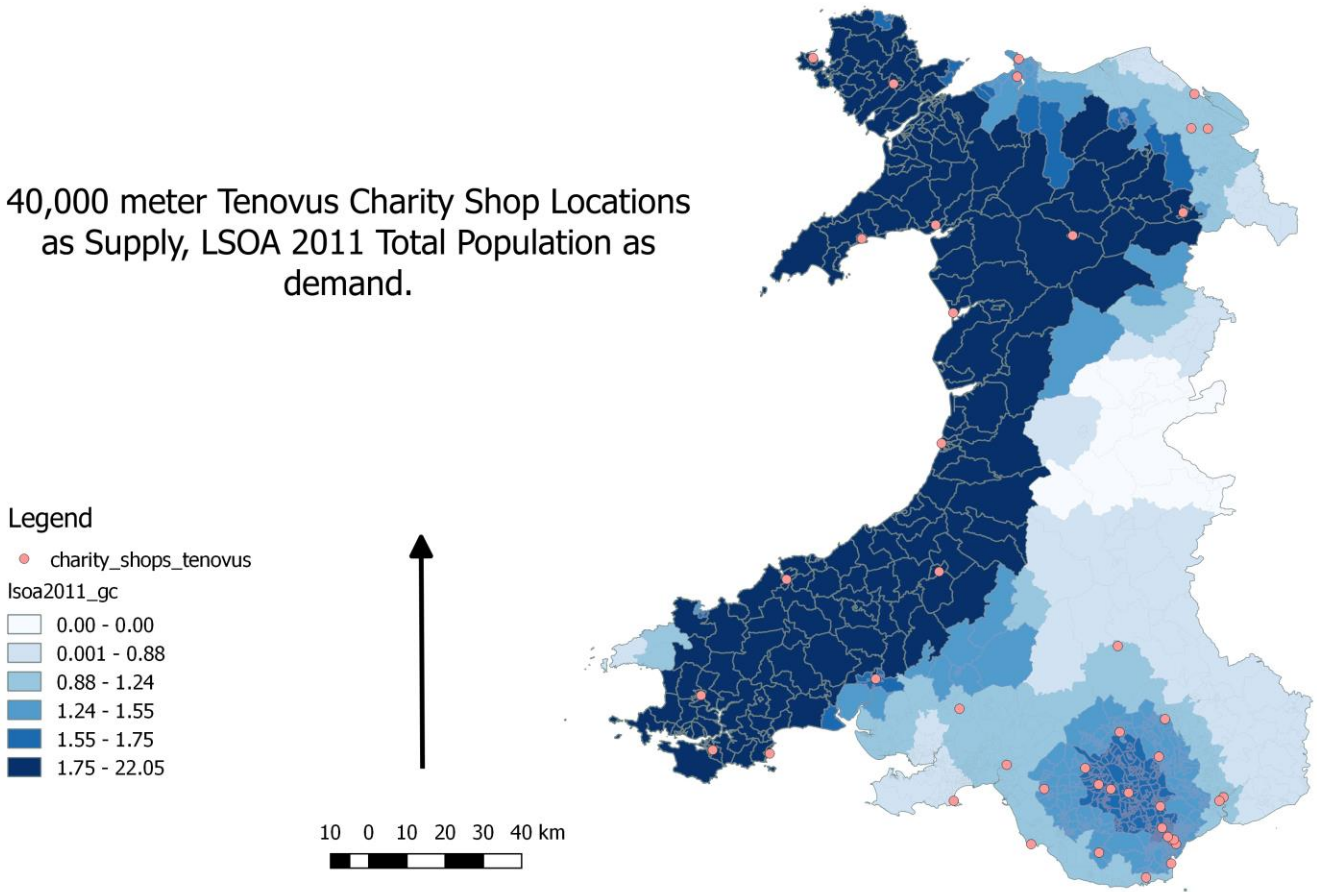

40,000 meter Tenovus Charity Shop Locations as Supply, LSOA 2011 Total Population as demand.
Legend
charity_shops_tenovus
lsoa2011_gc
0.00 - 0.00
0.001 - 0.88
0.88 - 1.24
1.24 - 1.55
1.55 - 1.75
1.75 - 22.05
10 0 10 20 30 40 km

Figure 5

Appendix D User Studies

D.1 User study 1 Documentation

2SFCA Tool user study

This tutorial walks the user through how to use the 2SFCA tool from opening it to creating a map using QGIS.

Open the tool from the file explorer, C:\2sfca.

Complete the boxes on the left-hand side – there is a lot of information required for the program to run.

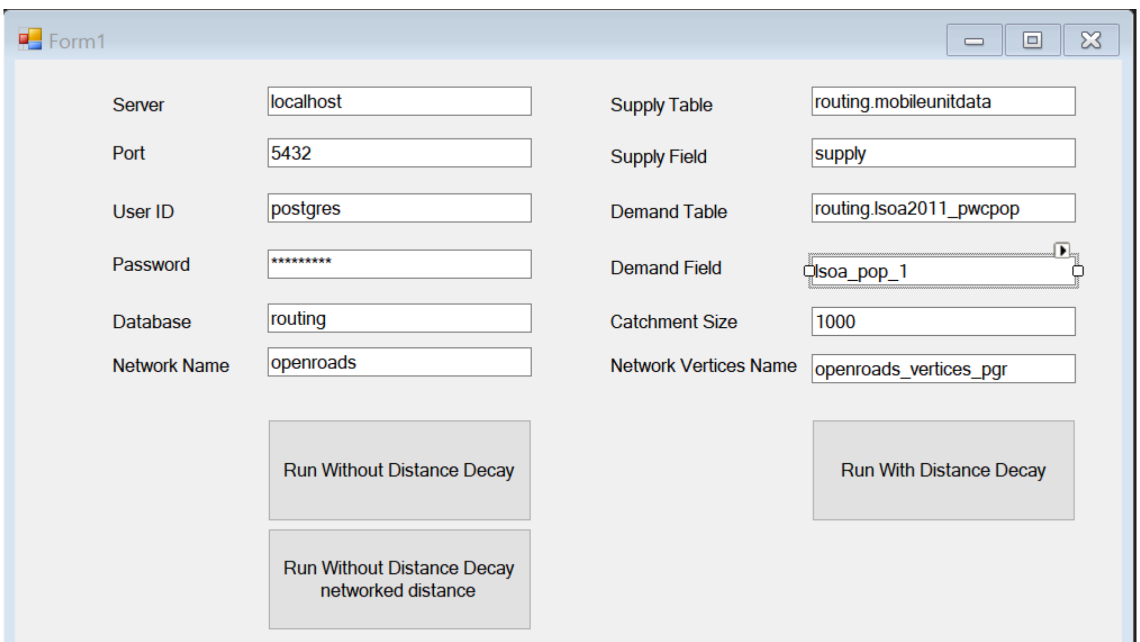


Figure 1

Complete the supply boxes on the right-hand side of the table – this is how the program knows what data to take for the computations. The supply table in the example is all of the Tenovus mobile unit data that is held in the database and the supply field is the particular field of interest within that table. This means that we are interested in the mobile units table and that the supply field within that table is where we are getting our supply value from.

Complete the demand boxes on the right-hand side of the table – in this example population weighted centroids are used at an LSOA level and we are using the population of each LSOA.

Catchment size – this is the size of the catchment that we are interested in. If you are looking at a large area like a country then a larger catchment is advised and if you are looking at smaller areas then a smaller catchment is better. It should be based on how far the demand side is willing to travel for the service, in the case of cancer care that can be quite far but for a pharmacist it may be a lot smaller. In this case try catchment size of 30,000 which is in metres.

Run the program – Initially run the without distance decay button. This runs a straight-line program like the one that was created in QGIS, it is quicker to run it this way.

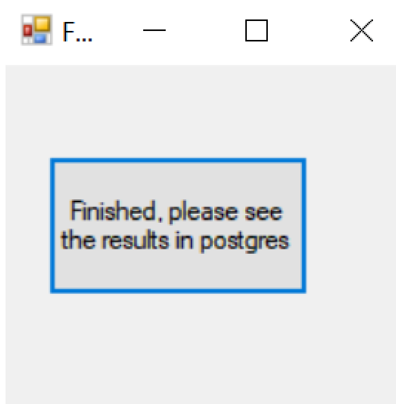


Figure 2

Close the program – Once you see this the process has been completed and we have our results. Click the button to exit the program.

Check the results – Open QGIS 2.14.14. In the browser panel on the left-hand side of the screen scroll down to the bottom and look at the PostGIS tables. Right click on the routing schema and click refresh which updates the tables.

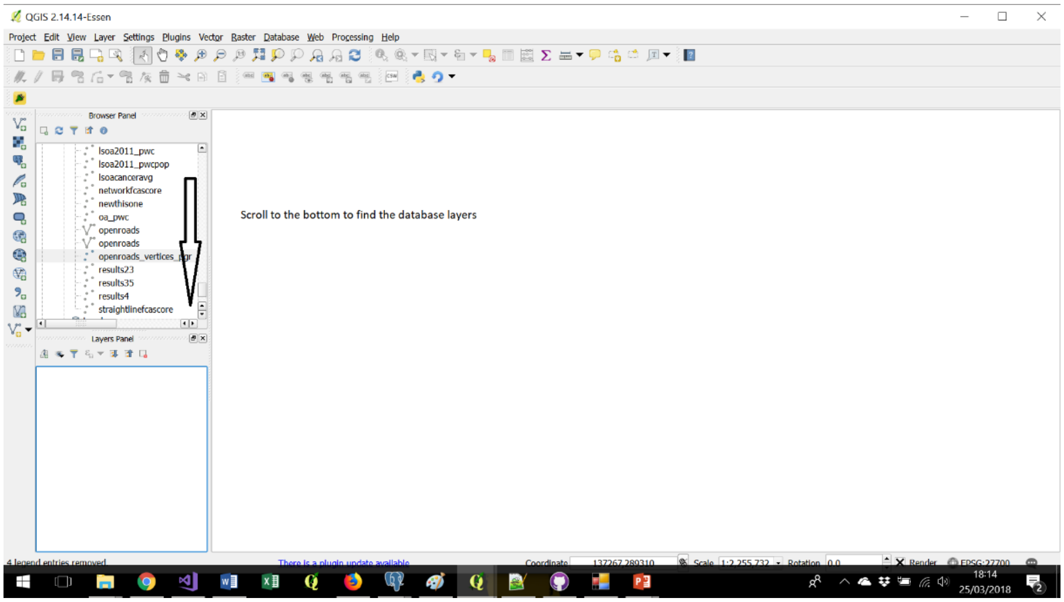


Figure 3

Open the LSOA2011_gc layer by double clicking on it, this layer has the boundaries of the LSOAs in Wales, do this twice so there are two open.

Open the straightlinefcascore table by double clicking on it. This is the layer which holds all of the scores. If you right click on it and select attribute table it is possible to see the raw data we are going to display. The column of interest is Sum which holds the 2SFCA scores.

	gid	code	demand	sum
0	1	W01001546	1	0.0076159...
1	2	W01001545	1	0.0076159...
2	3	W01001542	1	0.0066716...
3	4	W01001543	1	0.0057007...
4	5	W01001540	1	0.0087693...
5	6	W01001541	1	0.0087693...
6	7	W01001548	1	0.0034611...
7	8	W01001549	1	0.0057007...
8	9	W01000906	1	0.0103653...
9	10	W01000987	1	0.0091729...
10	11	W01000907	1	0.0103653...
11	12	W01000986	1	0.0103237...
12	13	W01000904	1	0.0103653...
13	14	W01000985	1	0.0103237...
14	15	W01000905	1	0.0103653...
15	16	W01000984	1	0.0091729...
16	17	W01000902	1	0.0103653...
17	18	W01000983	1	0.0103237...
18	19	W01000903	1	0.0103653...
19	20	W01000982	1	0.0091729...
20	21	W01000981	1	0.0091729...
21	22	W01000901	1	0.0084295...
22	23	W01000980	1	0.0091729...
23	24	W01000908	1	0.0073802...
24	25	W01000989	1	0.0091729...
25	26	W01000909	1	0.0073802...
26	27	W01001436	1	0.0097814...
27	28	W01001437	1	0.0097814...
28	29	W01001434	1	0.0097814...

Figure 4

Open the mobile unit data layer as you did before, locate it and double click it to add it to the screen.

Join – Now we need to join the tables, to do this double click on the LSOA2011_gc layer. This brings up a menu, from the menu select the joins tab. Complete the fields as below and click OK then apply. This is telling QGIS which fields to join the two tables on and the code is a unique reference. Eventually this is done automatically to save the user time.

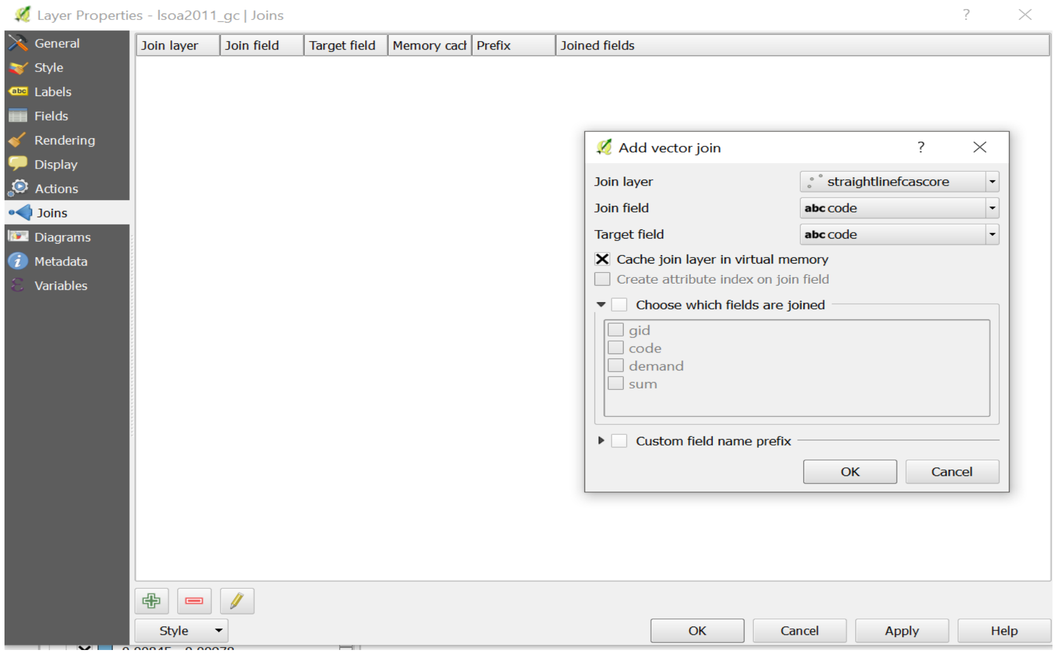


Figure 5

Map the data – All the data that is needed is in one location now, we finally need to visualise it in map form by accessing the style tab which is just above the join tab.

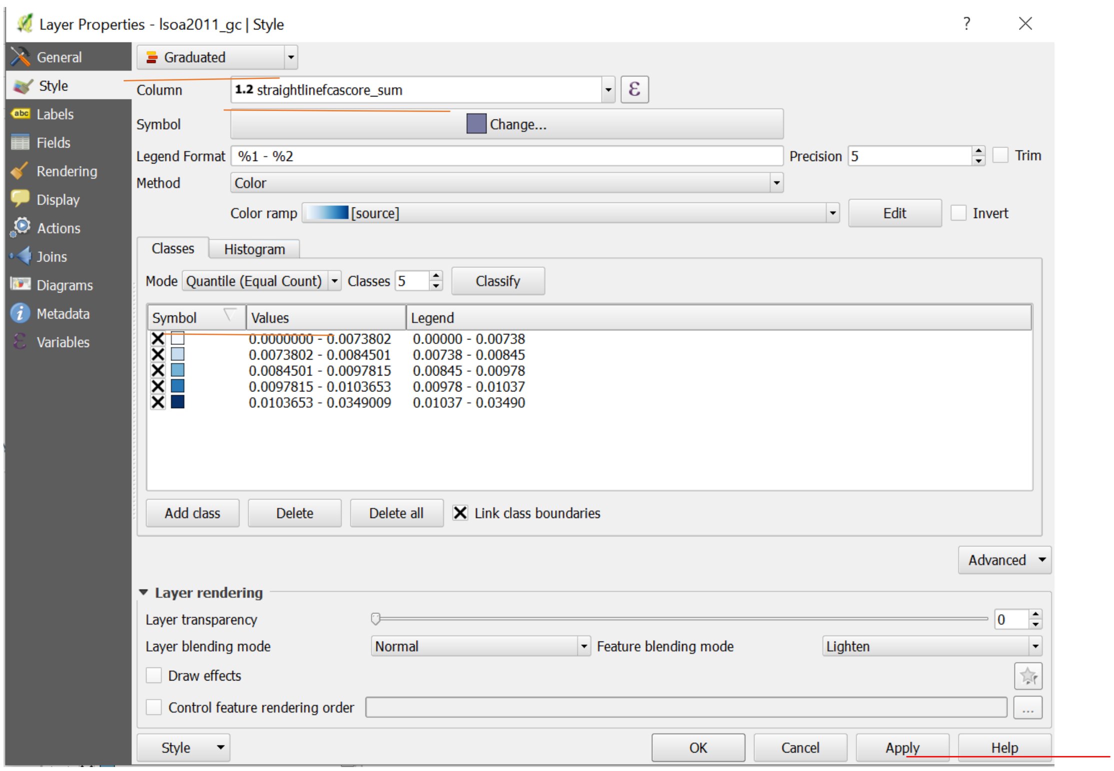


Figure 6

Graduate – as shown above there are many ways to visualise information using QGIS but for the best results with this type of data it is better to follow the above. First we select Graduated at the top, then we select the column that we want to visualise; in this case it is the straightlinefcascore_sum column. I like to use the blues colour ramp but there are many to experiment with, if you click on this option you can try and see which ones you think work best with the data. We then need to select the Quantile mode, these can be experimented with, and then press classify. The last thing we need to do on this page is to lighten the feature blending mode as it makes it easier to see the map. Once this is completed press the apply button and then OK.

Show original supply locations – drag the mobile unit data layer to the top and it should be possible to see the effect that each choir is having on its surroundings.

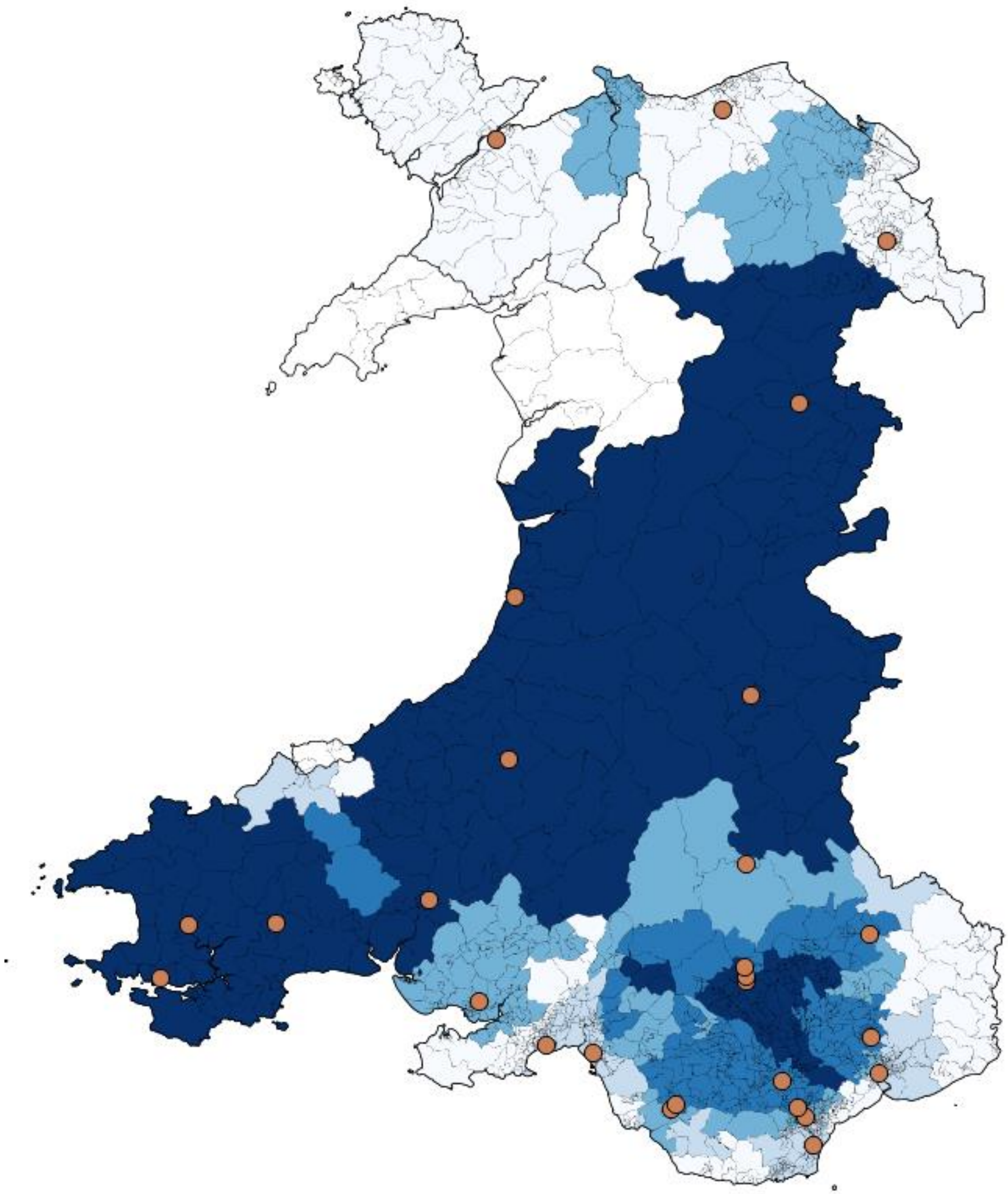


Figure 7

Double click on the other LSOA2011_gc layer, change the colour to white and lighten this so that there is a white background showing the remaining places in Wales.

Saving – Once this is done it is possible to save this information to come back to at a later date. Right click on the selected layer then you can save as and select a location on the machine to save this data.

D.2 User study 2 documentation

2SFCA Tool user study

This tutorial walks the user through how to use the 2SFCA tool from opening the tool and then running the 2SFCA calculations through opening QGIS and saving the map on the local system.

Open the tool from the file explorer, C:\2sfca.

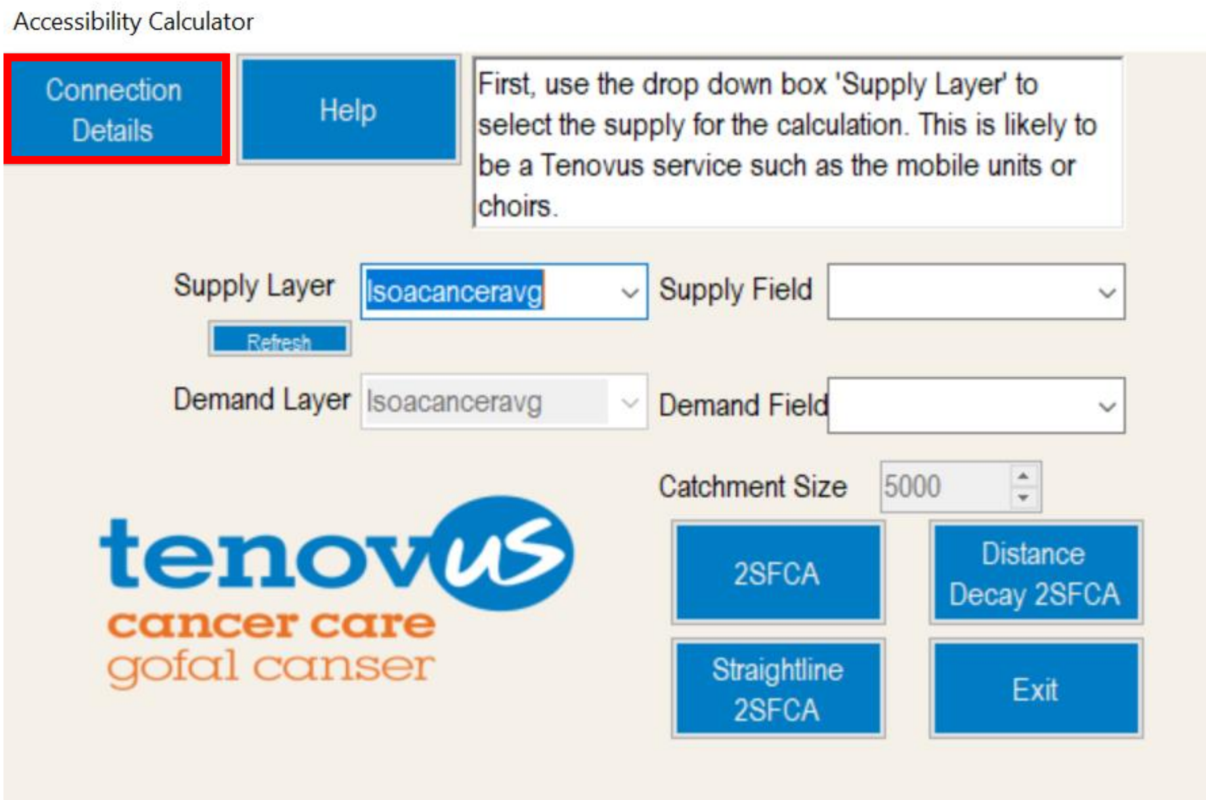


Figure 8

Click on the connection details box located in the top left-hand corner of the form. This displays the connection details and allows you to check that the connection to the database is working. These details can be changed but this will normally be pre-set and can be ignored.

Press the test connection button to ensure the connection is good.

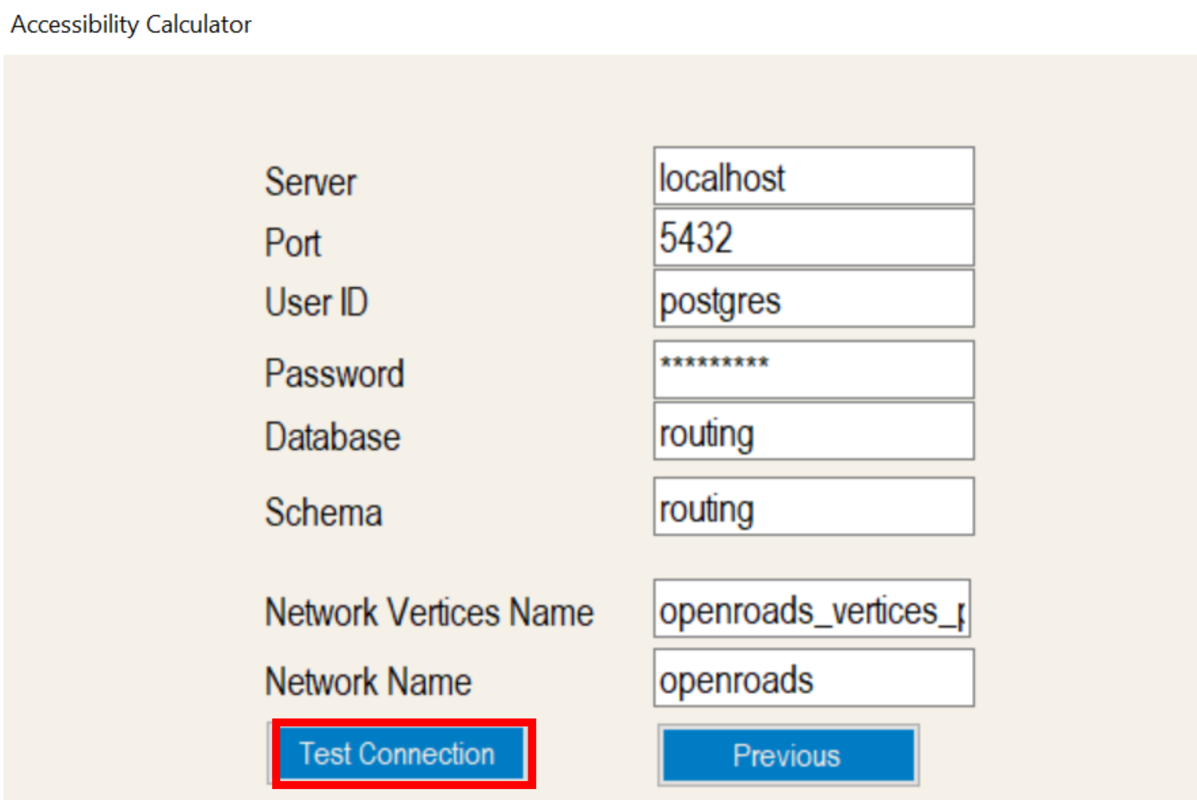


Figure 9

Press the previous button to get to the main screen.

The main screen has several options which need to be completed. The help button opens the full help document in text file, but the box has directions for what needs to be input at what time and if you hover the mouse over any button it will display a tip about what each button does.

Input the supply layer by selecting an option from the drop-down list. For this example, select ‘choir’. This will allow the computer to find all the data associated with choirs and will populate the next box.

Figure 10

The next box that needs to be completed is the supply field box, this box tells the program what will be used as a supply amount. In this case the option to select is 'supply' but any numerical unit could be used. For example, this could be the maximum capacity of a choir or the number of treatments that a mobile unit could perform in a day.

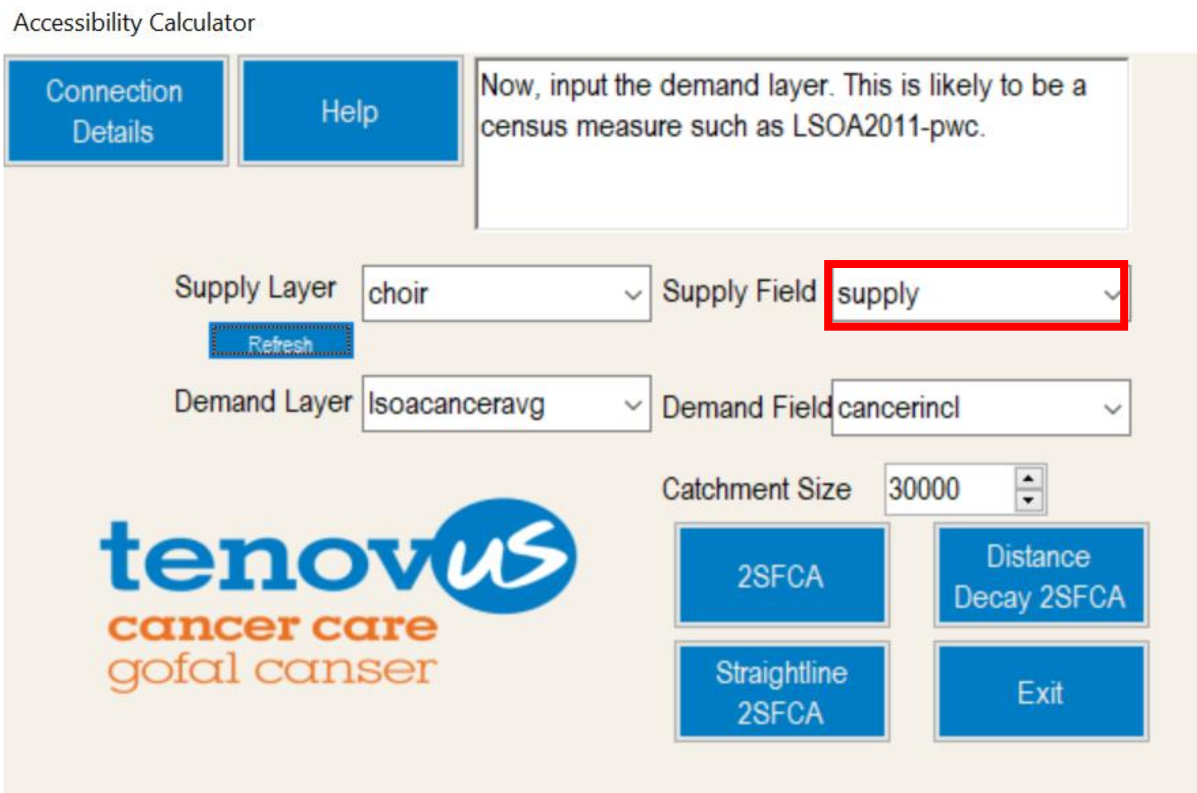


Figure 11

The demand layer box needs to have the 'lsoacanceravg' option selected. This is the LSOA Cancer Incidence (indirectly age-sex standardised) (number per 100,000) from the Welsh index of multiple deprivation. This could be anything which you use to identify people who may use your services but LSOA level or finer provides better results.

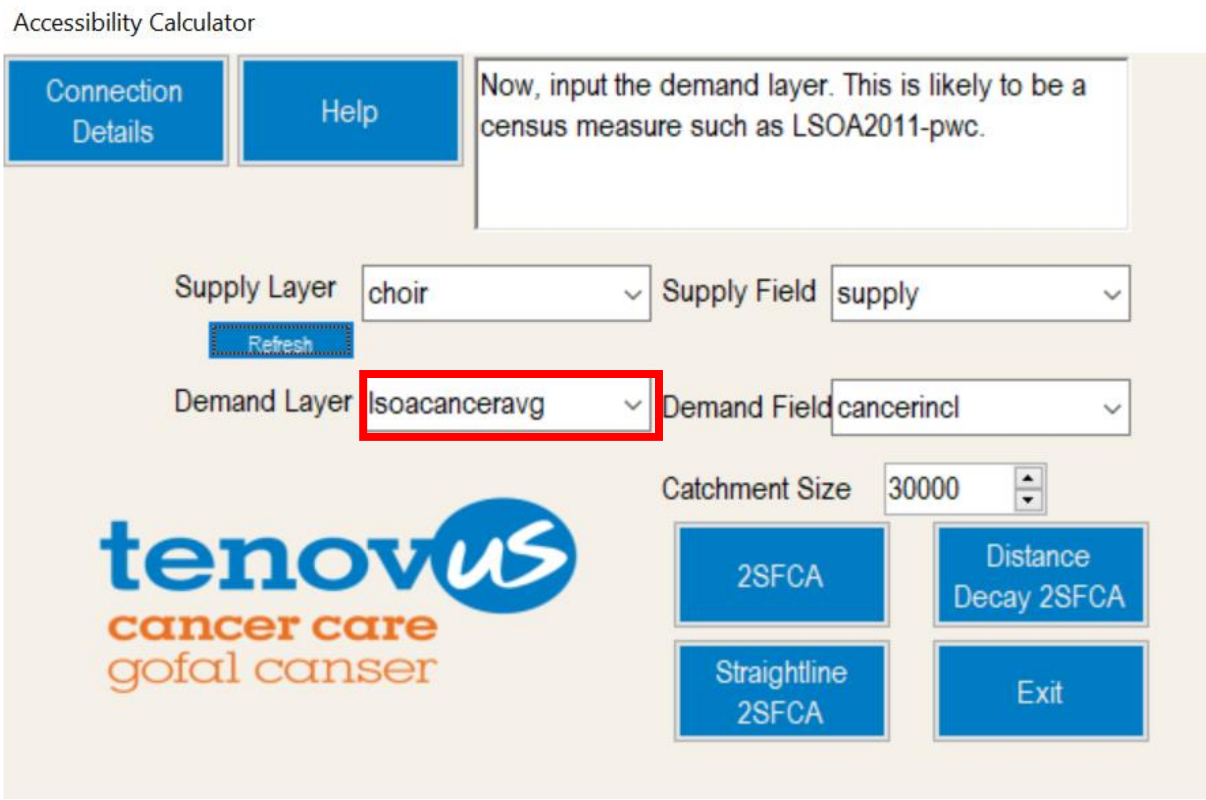


Figure 12

For the demand field the ‘cancerincl’ option needs to be selected, which is the actual number for each LSOA.

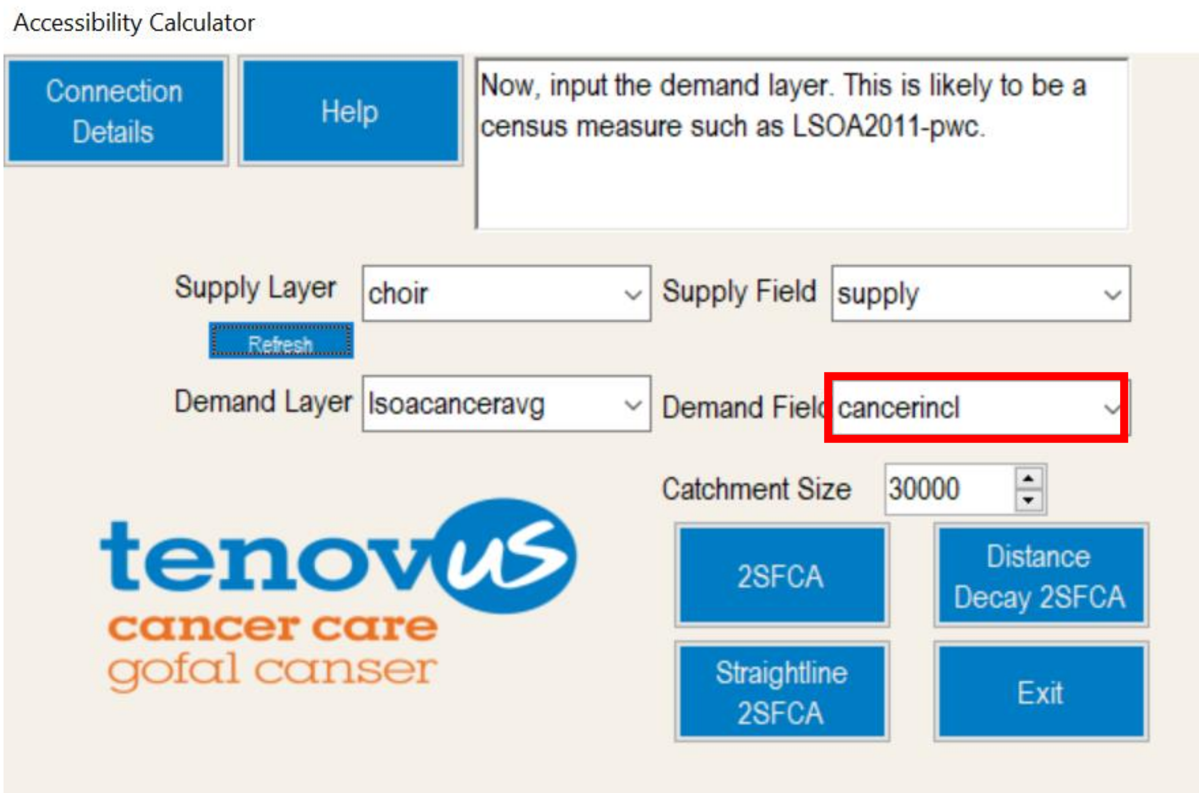


Figure 13

In this example, the program is using the Tenovus choir sites with cancer incidence rate. This should provide an overview of how well positioned the Tenovus choirs are, and highlight any potential gaps in coverage across Wales.

Catchment size – this is the size of the catchment that we are interested in. If you are looking at a large area like a country then a larger catchment is advised and if you are looking at smaller areas then a smaller catchment is better. It should be based on how far the demand side is willing to travel for the service, in the case of cancer care that can be quite far but for a pharmacist it may be a lot smaller. In this case try catchment size of 30,000 which is in metres.

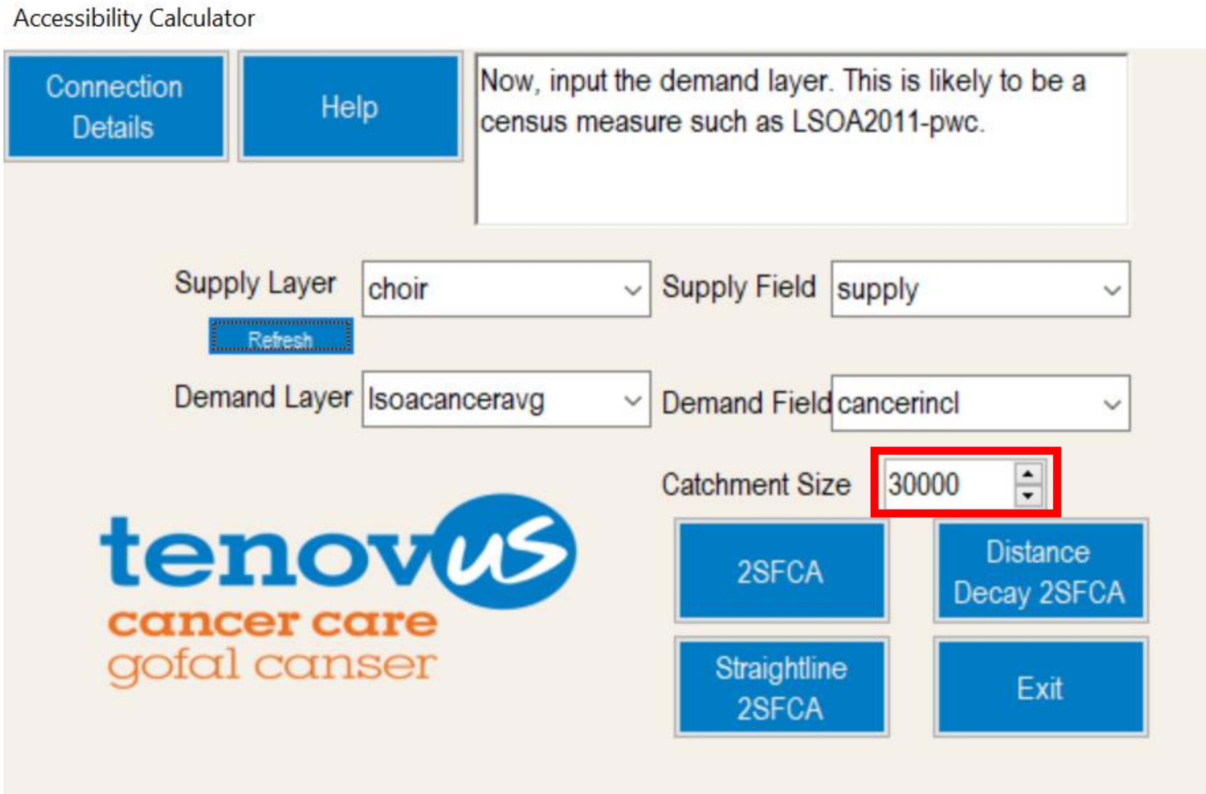


Figure 14

Run the program – Initially the standard ‘2SFCA’ calculation by pressing the 2SFCA button located at the bottom right of the screen. This uses the supply (choir) and the demand (cancer incidence) to better understand how accessible the choirs are and also highlight areas with poor accessibility.

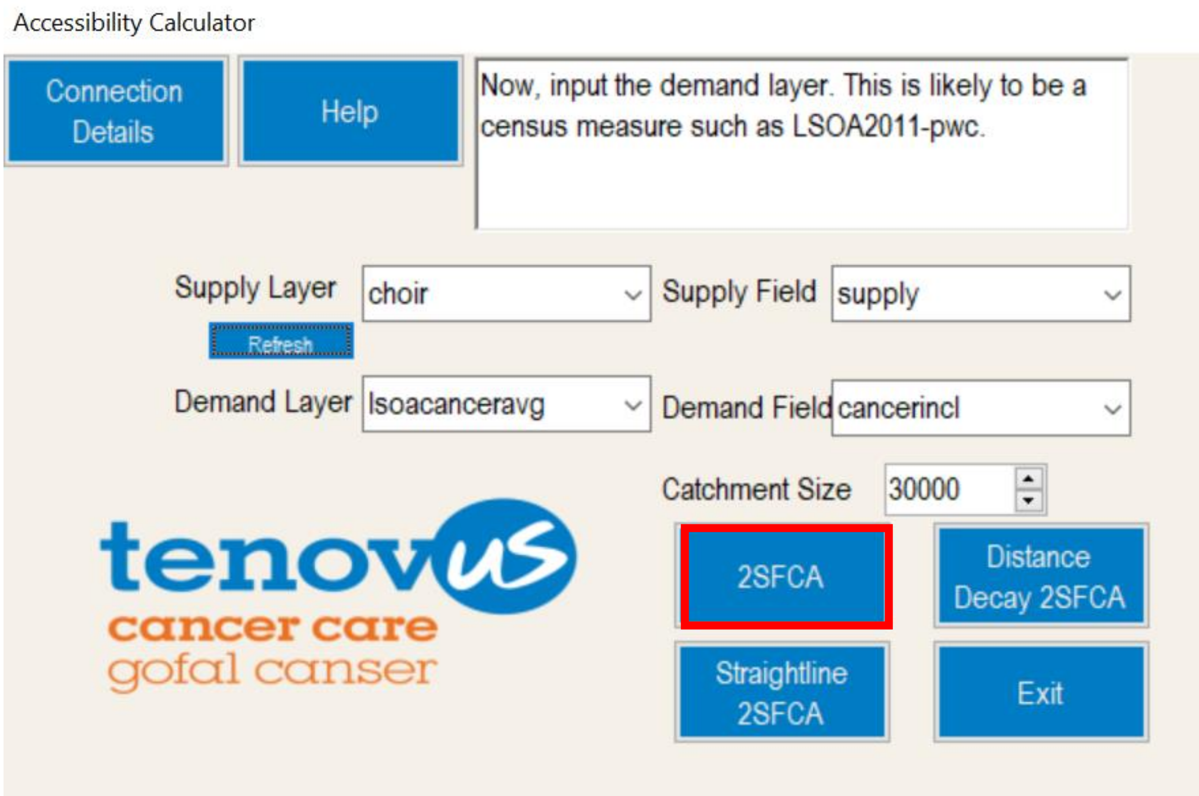


Figure 15

Once the program has completed the following message box will appear.

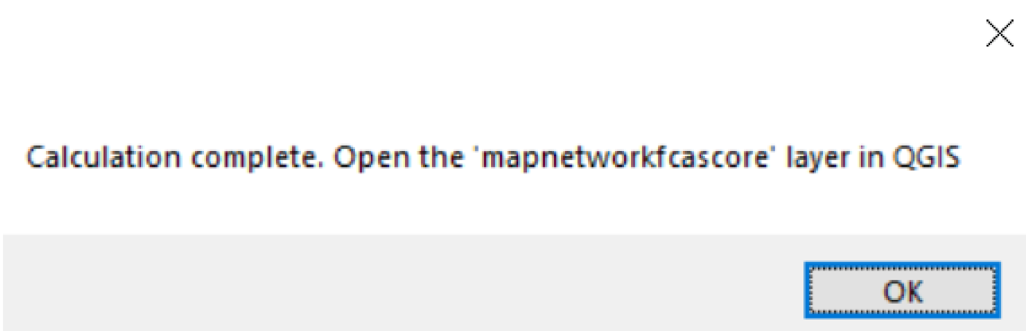


Figure 16

Close the program – Once you see that this the process has been completed we have our results. Click the button and then exit the program.

Check the results – Open QGIS 2.14.14. In the browser panel on the left-hand side of the screen scroll down to the bottom and look at the PostGIS tables. Right click on the routing schema and click refresh which will update the tables.

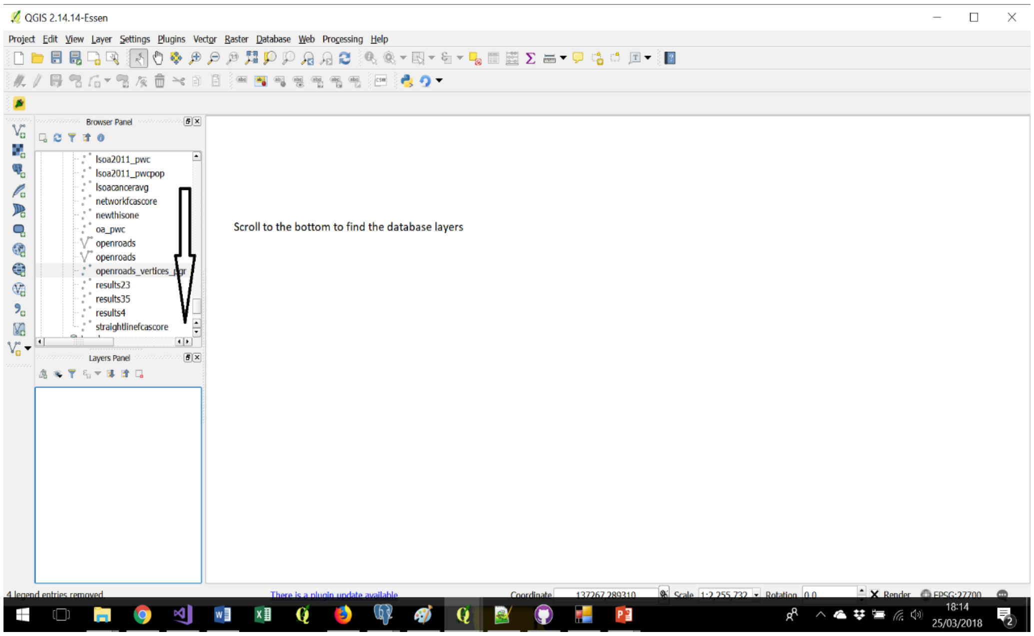


Figure 17

Open the LSOA2011_gc layer – do this by double clicking on it, this layer has the boundaries of the LSOAs in Wales. Double click on the layer once it is open and change the colour to white.

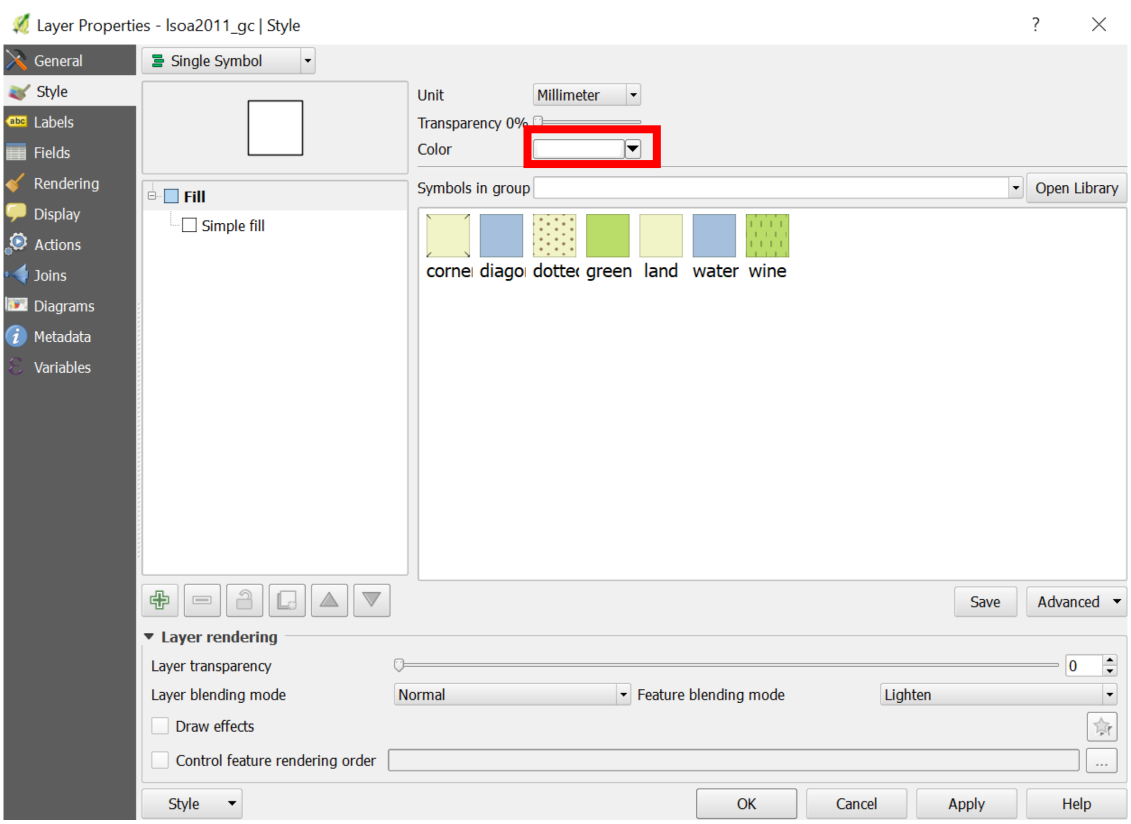


Figure 18

Open the mapnetworkfcascore table by double clicking on it. This is the layer which holds all of the scores, if you right click on it and select attribute table it is possible to see the raw data we are going to display. The column of interest is fcascore which holds the 2SFCA scores.

mapnetworkfcascore :: Features total: 507, filtered: 507, selected: 0

	code	gid	fcascore
0	W01001546	1527	0.0001739...
1	W01001545	1526	0.0001739...
2	W01000987	983	5.1923775...
3	W01000986	982	5.1923775...
4	W01000985	981	5.1923775...
5	W01000984	980	5.1923775...
6	W01000983	979	5.1923775...
7	W01000982	978	5.1923775...
8	W01000981	977	5.1923775...
9	W01000980	976	5.1923775...
10	W01000989	984	5.1923775...
11	W01000866	866	2.8619017...
12	W01000867	867	2.8619017...
13	W01000864	863	2.8619017...
14	W01000865	865	2.8619017...
15	W01000862	861	2.8619017...
16	W01000863	862	2.8619017...
17	W01000860	859	2.8619017...
18	W01000861	860	2.8619017...
19	W01000868	868	2.8619017...
20	W01000869	869	2.8619017...
21	W01000156	169	8.5738292...
22	W01000157	170	8.5738292...
23	W01000154	167	8.5738292...
24	W01000155	168	8.5738292...
25	W01000158	70	8.5738292...
26	W01000367	369	3.8117891...
27	W01000365	367	3.8117891...
28	W01000368	370	3.8117891...

Figure 19

Open the choir layer – do the same as you did before, locate it and double click it to add it to the screen.

Map the data – All the data that is needed is in one location now, we finally need to visualise it in map form. Double click on the mapnetworkfcascore layer, and click on the style tab.

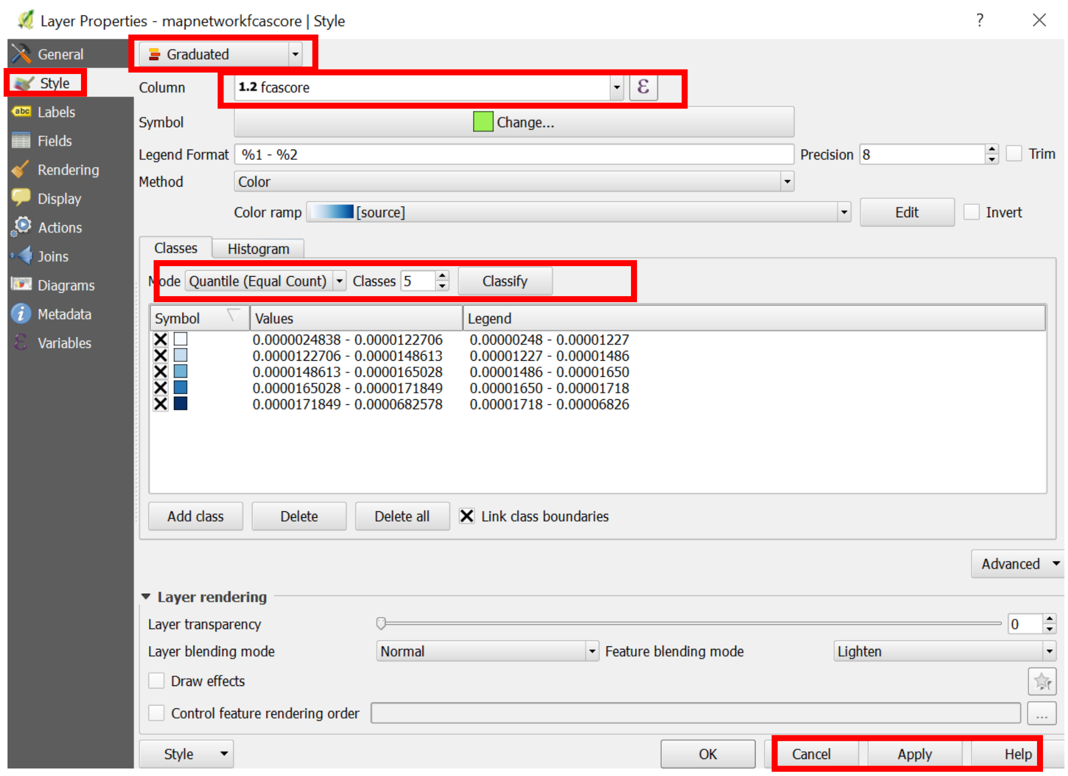


Figure 20

Graduate –there are many ways to visualise information using QGIS but for the best results with this type of data, it is better to follow the above. First, we select Graduated at the top, then we select the column that we want to visualise in this case it is the fcascore column. I like to use the blues colour ramp but there are many to experiment with, if you click on this option you can try and see which ones you think work best with the data. We then need to select the Quantile mode but again these can be experimented with and then press classify. The last thing we need to do on this page is to lighten the feature blending mode as it makes it easier to see the map. Once this is completed press the apply button and then OK.

Show original supply locations – drag the choir layer to the top and it should be possible to see the effect that each choir is having on its surrounding.

The map below shows what you should have, the darker the blue the better the accessibility. It is possible to see from the results that although the south east of Wales has many of the choirs there are still gaps where people have poor access to them. This may be as they are too far away or because they are sharing the choir with many other users.

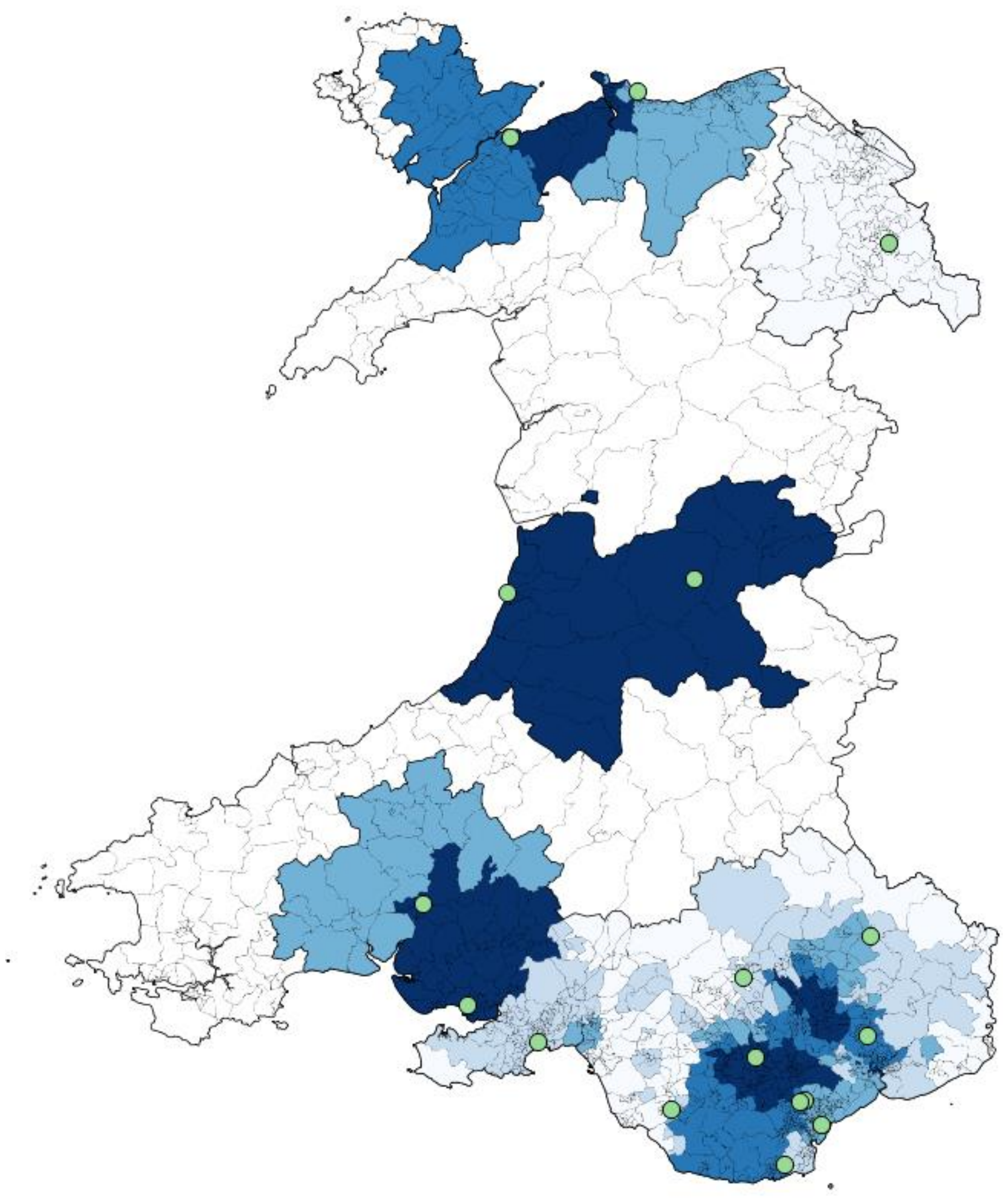


Figure 21

Saving – Once this is done it is possible to save this information to come back to at a later date. If you right click on the selected layer then you can save as and select a location on the machine to save this data. If you save the map on the desktop as 'yourname'_choircancermap it can be looked at later on.

Now we will use different data and create a proper map which can be shared.

Open the tool from the file explorer, C:\2sfca.

This time we will look at the mobile unit data and actual population of each LSOA.

Populate the boxes as you did previously with the supply layer as 'mobileunitdata' and the supply field as 'supply'. The demand layer is 'lsoa2011_pwcpop' and the demand field is 'lsoa_pop_1'. Finally, the size of the catchments will be 40,000 metres.

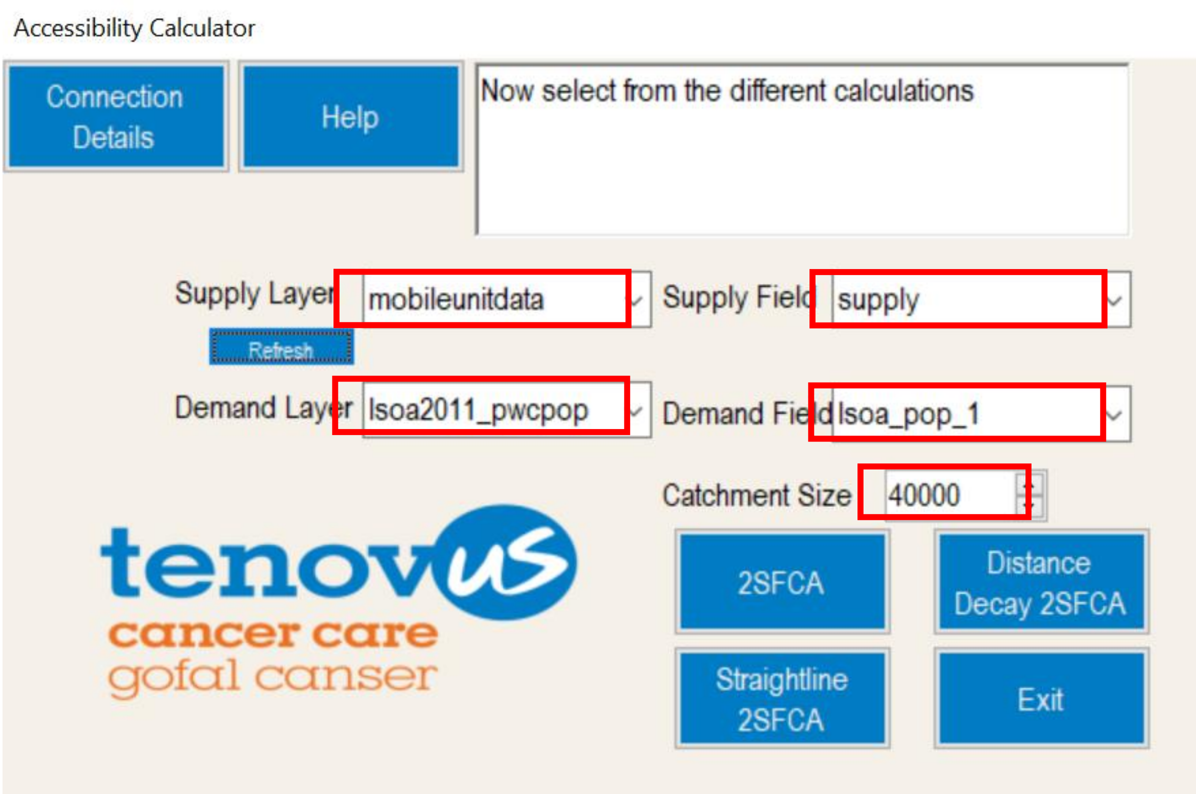


Figure 22

Run the 2SFCA calculation by pressing the 2SFCA button and exit the program once it has completed.

Open QGIS and open the mobile unit layer, lsoa2011_gc layer and mapnetworkfcascore layer.

Change the map as you did before to show the graduated scores. Double click the mapnetworkfcascore layer, opening the style tab and changing it to graduated. Then, change the mode to quantile and lighten the colours.

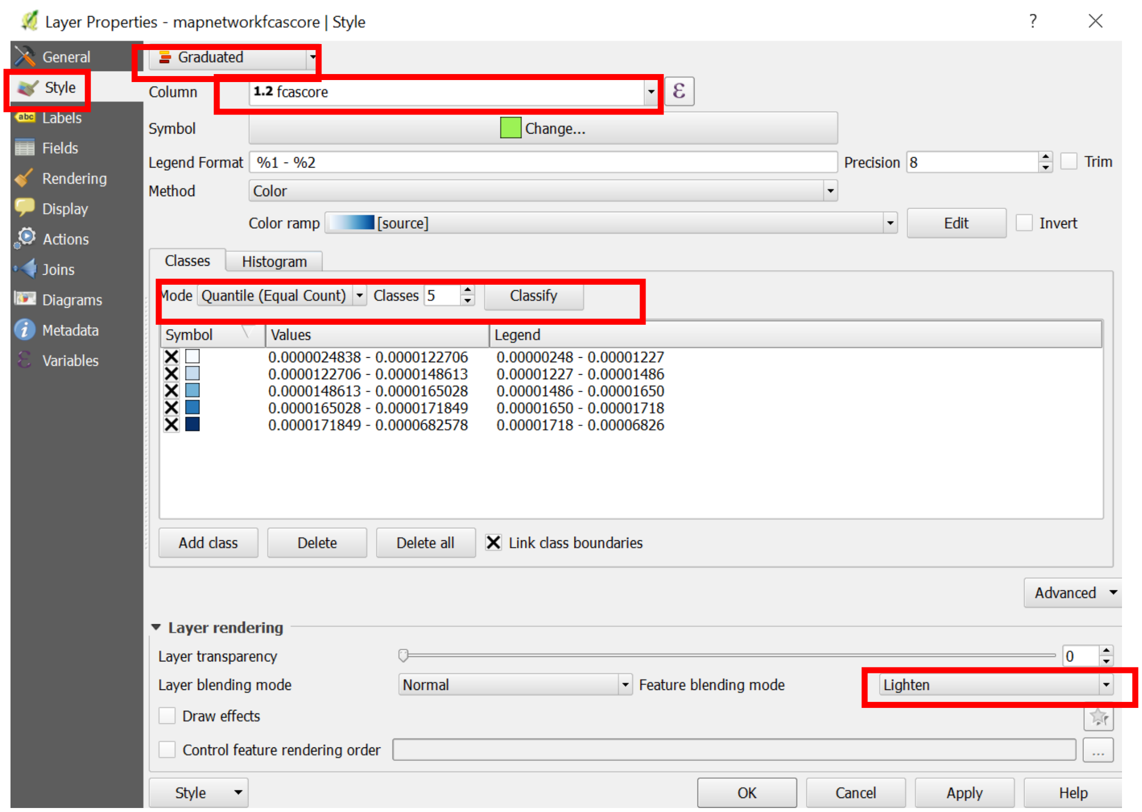


Figure 23

This will provide a map like last time but it does not look very professional. First, open the print composer in the project dropdown. It will ask you to save the project and this should be a sensible name which describes the data used such as, ‘population mobile unit 40km’.

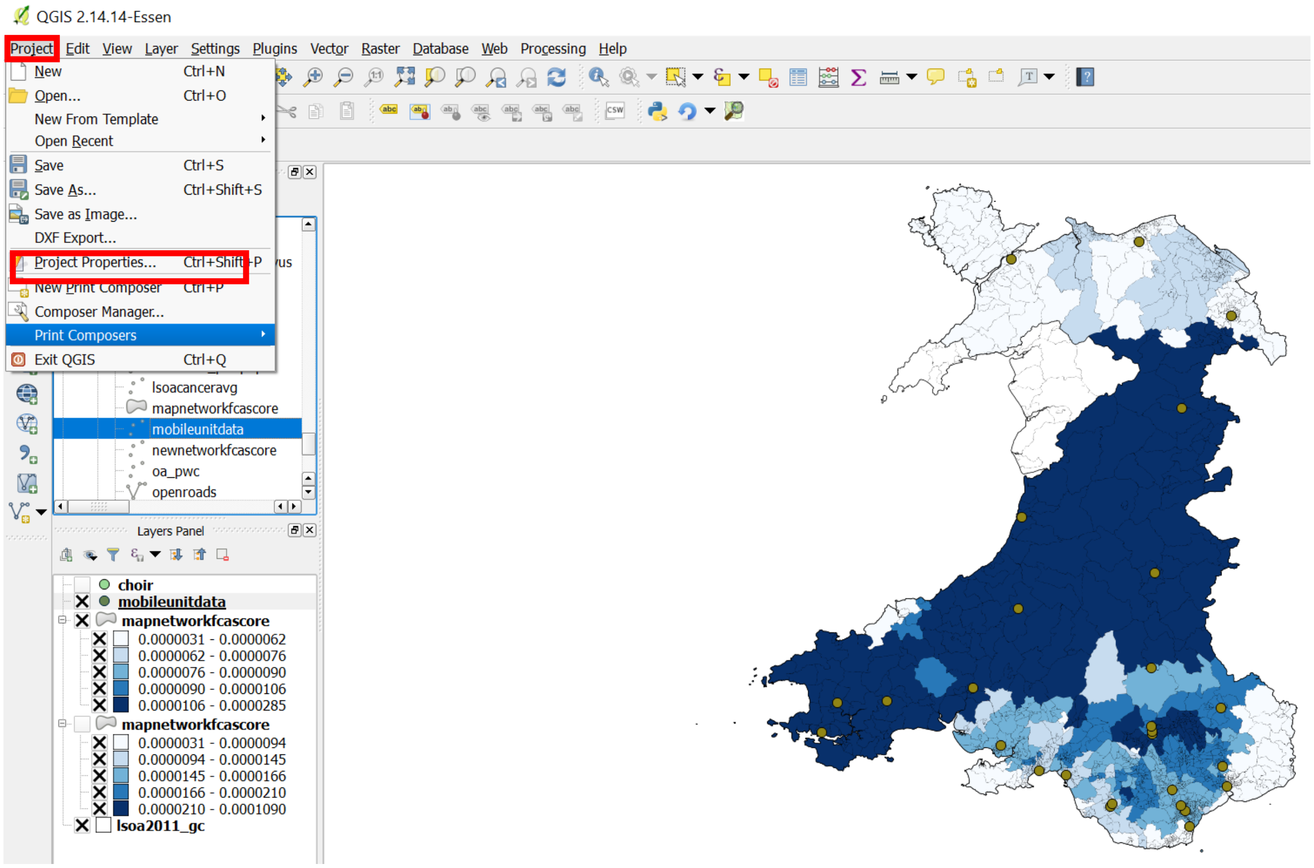


Figure 24

Once in the print composer the map needs to be added. To do this use the 5th button down on the right-hand side. You need to click in the top left hand corner and drag the box to cover the whole canvas.

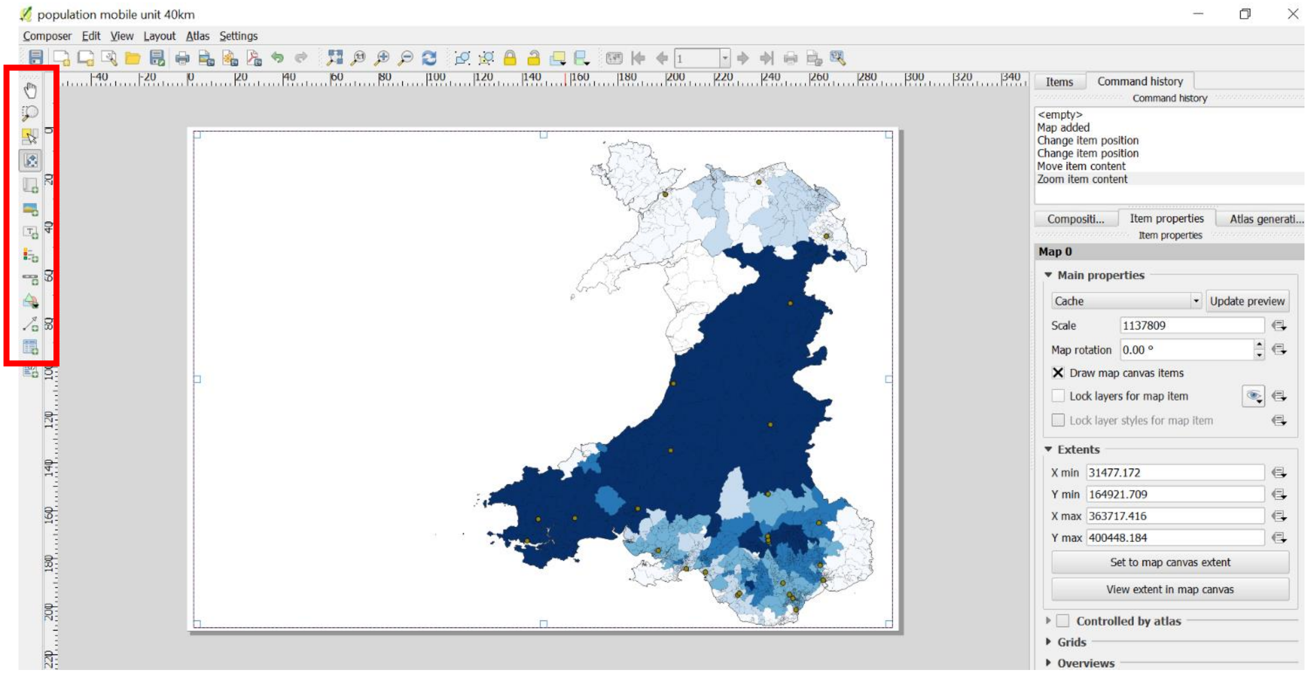


Figure 25

Once the map shows use the 4th button on the right-hand side to scale the map to fit across the right side of the panel.

A title of the map needs to be added as well, I have called mine 40km networked 2SFCA of Tenovus mobile units and 2011 census population. The results are displayed using a 5-classification quantile scheme. This needs to be descriptive and describe the data being used. This is accessed by using the add label button which is the 7th down on the left-hand side. Once selected it is possible to enter the title on the right side of the screen and by clicking on font we can change it to Arial and make it bigger.

A legend needs to be added, this takes the information from the classification that was done in QGIS but it can be edited to appear more nicely. This is the 8th button down on the left.

Finally, a scale bar and a north arrow need to be added both using their respective buttons (9 and 11). These can be placed anywhere but it is better to make them smaller and discrete.

There is a lot of customisation that can be done at every stage during this process and it is beyond the scope of this project to explore it all, but there are many great tutorial videos on YouTube which can assist you in making the maps look great.

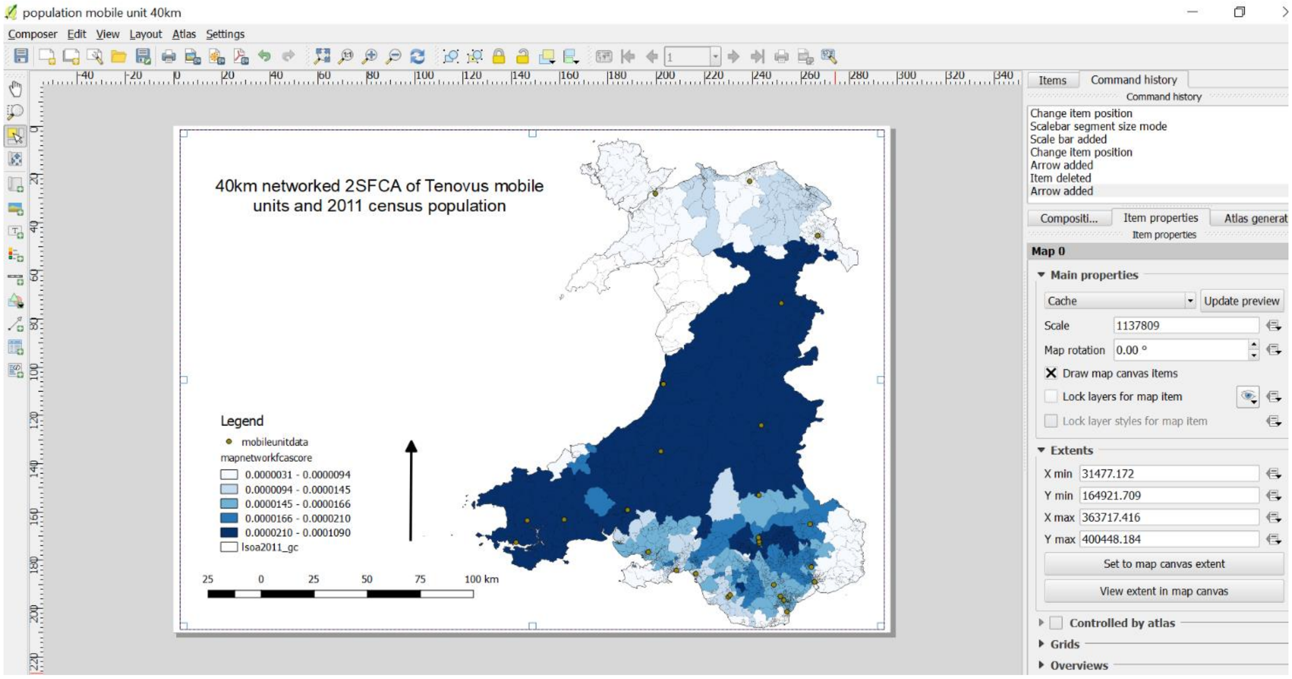


Figure 26

The image needs to be saved. This is best done using the PDF exporter and this is found in the composer tab. Save the PDF to desktop and call it 'yourname_40kmmobileunitmap'.

These are the basics of what this tool will allow you to do, and QGIS has a great deal of functionality which may be able to assist you in other parts of your job. It has a great deal of documentation and tutorials available freely online and if you plan to utilise this tool regularly it may be worth exploring in more detail.

D.3 User Experience Study 1 Questionnaire

User Experience Study 1 Questionnaire

1. What is your Job Title?

2. What is your age?

- ◯ 18 to 21
- ◯ 22-30
- ◯ 31-40
- ◯ 41-50
- ◯ 51-60
- ◯ 61+

3. How IT literate are you?

- ◯ Rarely use IT even for emails
- ◯ use It to browse the internet and look at emails
- ◯ use IT for work and use different types of programs (Word, Excel, Powerpoint)
- ◯ use advanced IT programs often (databases, etc)
- ◯ use advanced IT programs and can create content (programming and database management)

4. How did you find completing the task?

- ◯ Very easy
- ◯ Easy
- ◯ Moderate difficulty
- ◯ Difficult
- ◯ Very difficult

5. Did the tool feel familiar to other programs that you use?

- ◯ Very familiar
- ◯ Familiar
- ◯ Unfamiliar

6. Did you find the tool intuitive?

- ◯ Very intuitive
- ◯ Intuitive
- ◯ Not intuitive

7. Did you find creating the map difficult?

- ◯ Ver difficult
- ◯ Difficult
- ◯ OK
- ◯ Easy
- ◯ Very easy

8. Could you complete the task again without assistance?

- ◯ Yes
- ◯ Some of it
- ◯ Most of it
- ◯ No

9. Do you ever use database systems?

- ◯ Often
- ◯ Rarely
- ◯ Never

10. Do you think this would be better if it was online?

- ◯ Yes
- ◯ Unsure
- ◯ No

11. Would the additional functionality of QGIS be beneficial to you (creating maps and visualising spatial data)?

- ◯ Yes
- ◯ Unsure
- ◯ No

12. Did you understand the language used?

- ◯ None of it
- ◯ A little bit
- ◯ Most of it
- ◯ All of it

13. Was the tool nice to look at?

- ◯ Very nice
- ◯ Nice
- ◯ Acceptable
- ◯ Could be nicer
- ◯ Very unpleasant

14. Was the tool nice to use?

- ◯ Very nice
- ◯ Nice
- ◯ Acceptable
- ◯ Could be nicer
- ◯ Unpleasant

D.4 User Experience Study 1 Questionnaire results

Surveyplanet.com was used to create an online questionnaire. 14 questions were asked giving a good understanding of how the tool is viewed, and the general IT literacy of the Tenovus staff.

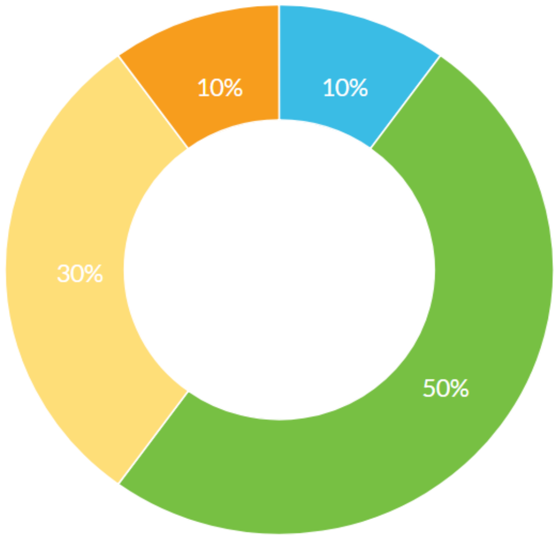
The questions and responses are as follows:

What is your job title?
PA, Head of Care, Marketing Manager, Research Officer, Prevention Lead, Prevention Campaign Advisor, Data Analyst, Choir Leader, Research Engagement Officer and Office Administrator.

The participant's positions within Tenovus varied from lower level administration assistants to Head of department. It is important to understand the variety of potential users and their roles within the organisation.

What is your age?
Four participants from 22-30 years old, four from 31-40 years old and two between 51-60 years old.

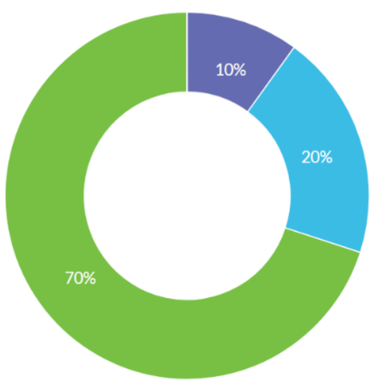
How IT literate are you?



Choices	Totals	Ratings
1 - rarely use IT even for emails	0	0
2 - use IT to look at emails and browse the internet	1	2
3 - use IT for work and use different types of programs (Word, Excel, Powerpoint)	5	15
4 - use advanced IT programs often (databases etc)	3	12
5 - use advanced programs and can create content (programming and database management)	1	5

Figure 27

How did you find completing the task?



Choices	Totals
Very easy	1
Easy	2
Moderate difficulty	7
Difficult	0
Very Difficult	0

Figure 28

Did the tool feel familiar to other programs that you use?

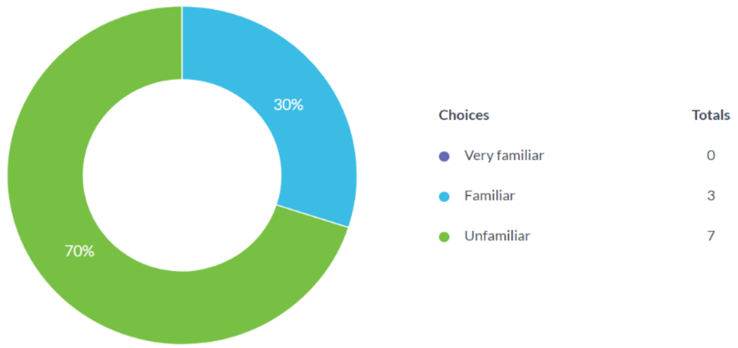


Figure 29

Did you find the tool intuitive?

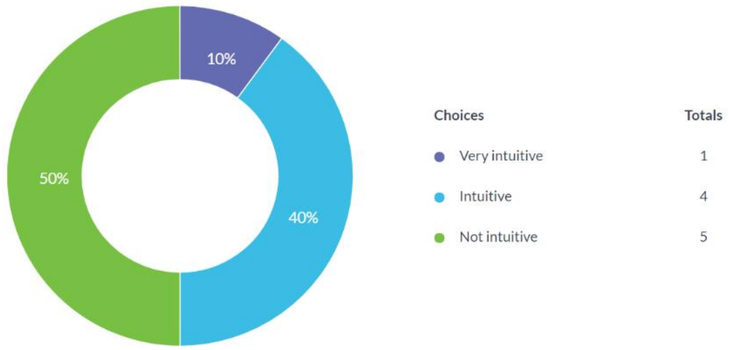


Figure 30

Did you find creating the map difficult?

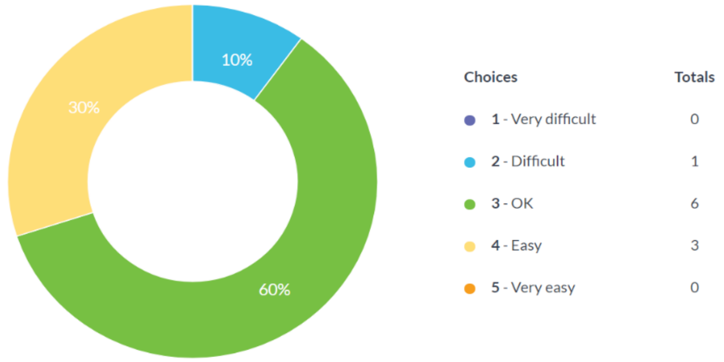


Figure 31

Could you complete the task again without assistance?

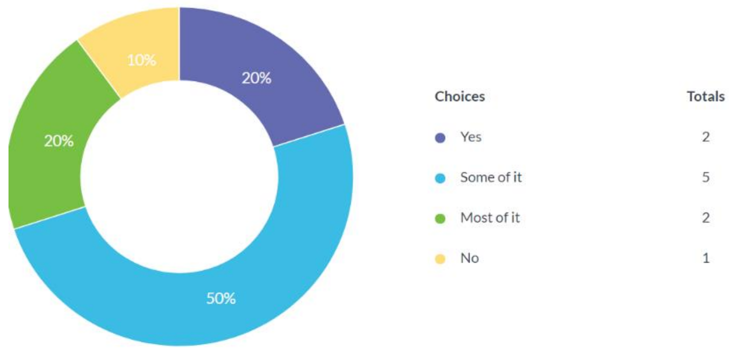


Figure 32

Do you ever use database systems?

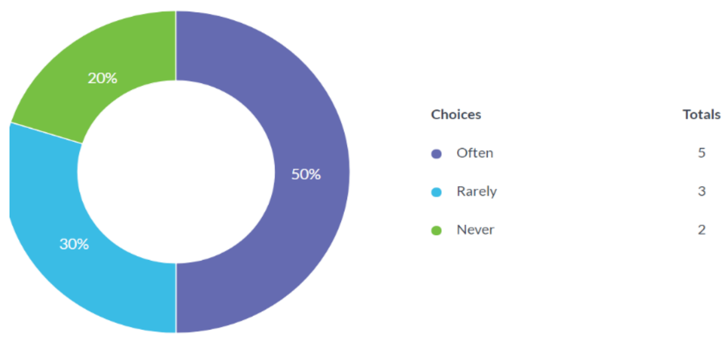


Figure 33

Do you think this would be better if it was online?

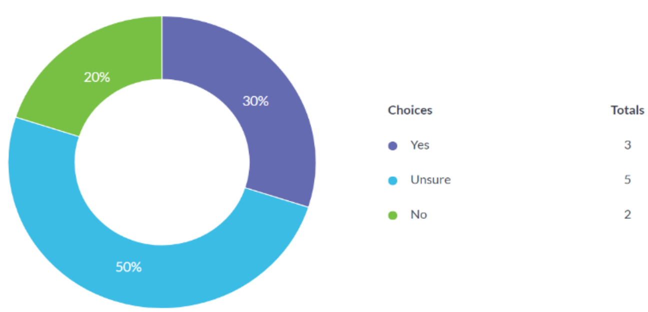


Figure 34

Would the additional functionality of QGIS be beneficial to you (creating maps and visualising spatial data)?

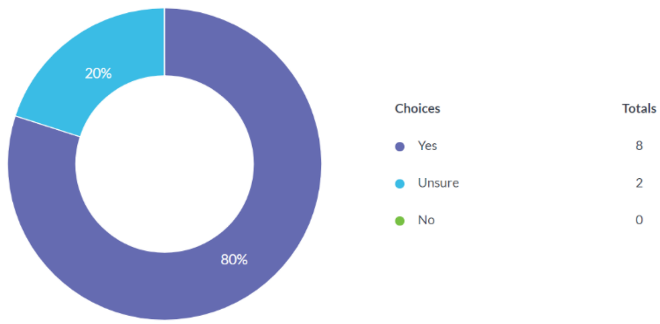


Figure 35

Did you understand the language used?

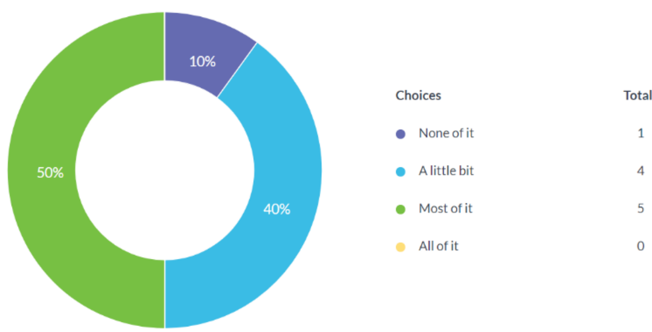


Figure 36

Was the tool nice to look at?

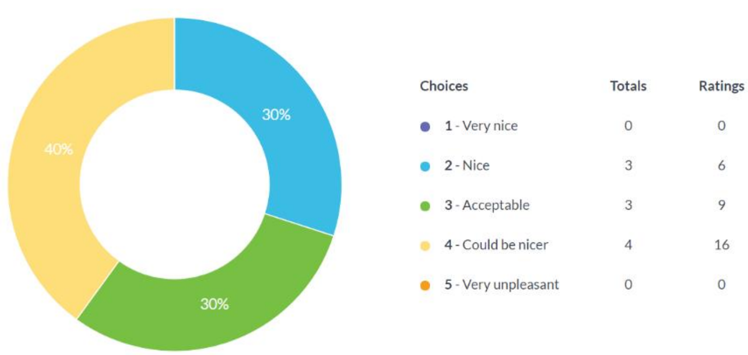


Figure 37

Was the tool nice to use?

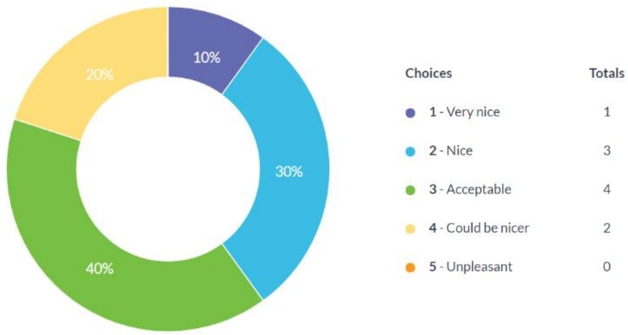


Figure 38

D.5 User Experience Study 2 Questionnaire

User Experience Study 2 Questionnaire

1. What is your job title?

2. What is your age?

- 18-21
- 22-30
- 31-40
- 41-50
- 51-60
- 61+

3. How IT literate are you?

- Rarely use IT even for emails
- Use IT to look at the emails and browse the internet
- Use IT for work and use different types of programs (Word, Excel, Powerpoint)
- Use advanced IT programs often (databases, etc)
- Use advanced programmes and can create content (programming and database management)

4. How did you find completing the task?

- Very easy
- Easy
- Moderate difficulty
- Difficult
- Very difficult

5. If you found the task difficult please expand on the parts of the process that were difficult or confusing?

6. Did the tool feel familiar to other programs that you use?

- ◯ Very familiar
- ◯ Familiar
- ◯ Unfamiliar

7. Did you find the tool intuitive?

- ◯ Very intuitive
- ◯ Intuitive
- ◯ Not intuitive

8. Did you find creating the map difficult?

- ◯ Very difficult
- ◯ Difficult
- ◯ OK
- ◯ Easy
- ◯ Very easy

9. Please provide any additional comments on how you found QGIS

10. Could you complete the task again without assistance?

- ◯ No
- ◯ Some of it
- ◯ Most of it
- ◯ Yes

11. Do you ever use database systems?

- ◯ Often
- ◯ Rarely
- ◯ Never

12. Would the additional functionality of QGIS be beneficial to you (creating maps and visualising spatial data)?

- ◯ Yes
- ◯ Unsure
- ◯ No

13. Could you run the program again with a different set of inputs?

- ◯ Yes
- ◯ Maybe
- ◯ No

14. Did you understand the language used?

- ◯ None of it
- ◯ A little bit
- ◯ Most of it
- ◯ All of it

15. Was the tool nice to look at?

- ◯ Very nice
- ◯ Nice
- ◯ Acceptable
- ◯ Could be nicer
- ◯ Very unpleasant

16. Was the tool nice to use?

- ◯ Very nice
- ◯ Nice
- ◯ Acceptable
- ◯ Could be nicer
- ◯ Unpleasant

17. If the tool was not nice to look at or easy to use please expand on the parts that could be improved.

18. Have the changes to the tool made it easier for you to use?

- ◯ A lot
- ◯ Yes
- ◯ Not really
- ◯ None at all

19. Does the tool fit in with the Tenovus brand?

- ◯ Fits the branding
- ◯ Not really
- ◯ Not at all

20. Please describe a task you might use the tool to complete

D.6 User Experience Study 2 Questionnaire results

Surveyplanet.com was used to create an online questionnaire, 20 questions were asked and this gives a good understanding of how the tool is viewed and the general IT literacy of the Tenovus staff.

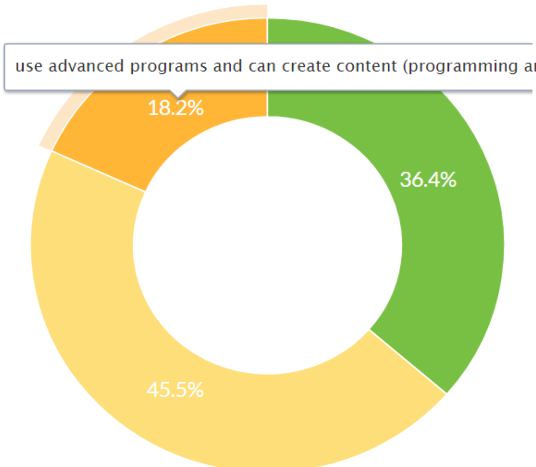
The questions asked were:

What is your job title?
PA, Head of Wellbeing, Head of Research, Marketing Manager, Research Officer, Prevention Campaign Advisor, Data Analyst, Choir Leader, Research Engagement Officer, Office Administrator and Head of Process Improvement.

This set of participants was similar to those in UX study 1. A wide range of positions are represented.

What is your age?
Four participants from 22-30 years old, four from 31-40 years old one each between 41-50, 51-60 and 60+ years old.

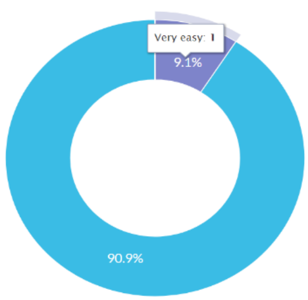
How IT literate are you?



Choices	Totals	Ratings
1 - rarely use IT even for emails	0	0
2 - use IT to look at emails and browse the internet	0	0
3 - use IT for work and use different types of programs (Word, Excel, Powerpoint)	4	12
4 - use advanced IT programs often (databases etc)	5	20
5 - use advanced programs and can create content (programming and database management)	2	10

Figure 39

How did you find completing the task?



Choices	Totals
Very easy	1
Easy	10
Moderate difficulty	0
Difficult	0
Very difficult	0

Figure 40

If you found the task difficult please expand on the parts of the process that were difficult or confusing?

A couple of occasions I might press the wrong button. If this was the case, I would need guidance as to how to get back to where I was originally to get on the right path.

Did the tool feel familiar to other programs that you use?

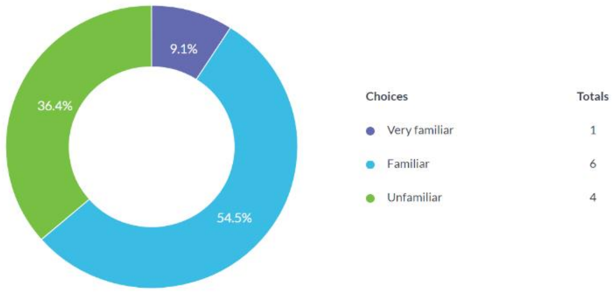


Figure 41

Did you find the tool intuitive?

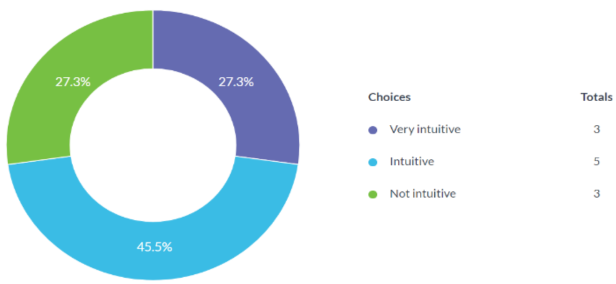


Figure 42

Did you find creating the map difficult?

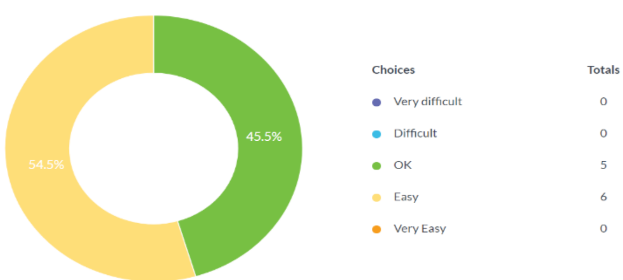


Figure 43

Please provide any additional comments on how you found QGIS

- Impressive and easy to use
- Generally good. Might be worth naming the fields, and clearing up the UI just to make it a little more intuitive.
- I found this really straightforward and interesting, I can see how a number of services would use it, the instructions were clear and easy to follow.
- Good program, which I only found slightly more difficult as I had never seen or used it before, but the functionality and user-friendliness of it was clear to see. I don't think it would be long before I would feel comfortable to use and manipulate it in whatever way I needed.
- I found it easy to use with the instructions and once used a few times it should be easy. It would be difficult to get to grips with without the instructions as it isn't like a program I use regularly.
- Useful tool for providing visual examples of geographic analytical data.
- Overall, I think it was excellent. I understood what I was doing, how and why. My only concern would be, as said previously, understanding what I might do to rectify a mistake if I was so reliant on the manual for the pathway to know what to do. A bit more knowledge and a lot more practice should overcome this.
- Once you're familiar with the options you need to use it's fine.

Could you complete the task again without assistance?

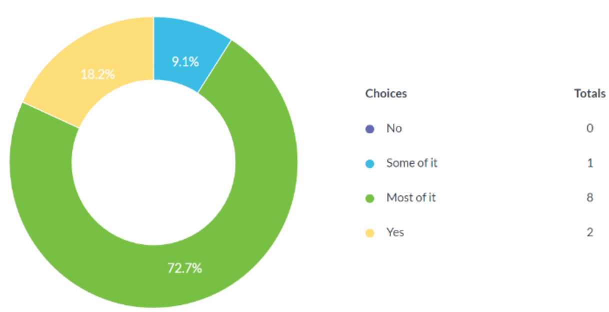


Figure 44

Do you ever use database systems?

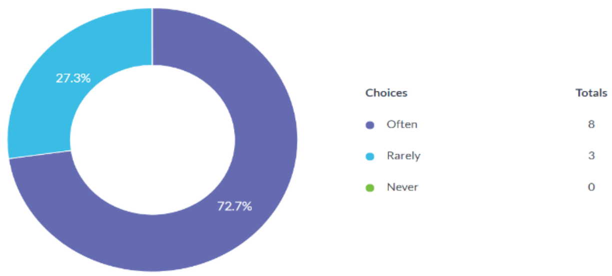


Figure 45

Would the additional functionality of QGIS be beneficial to you (creating maps and visualising spatial data)?

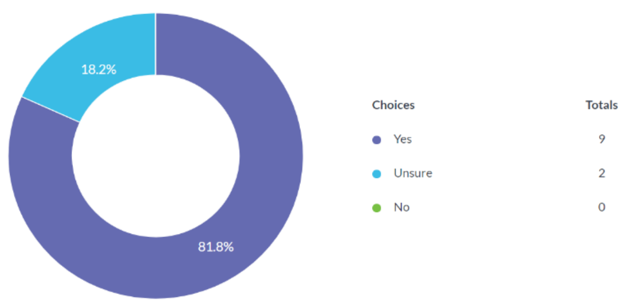


Figure 46

Could you run the program again with a different set of inputs?

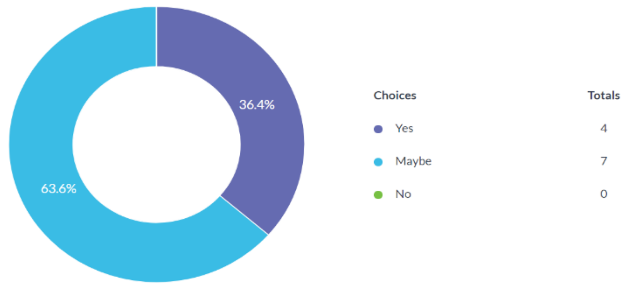


Figure 47

Did you understand the language used?

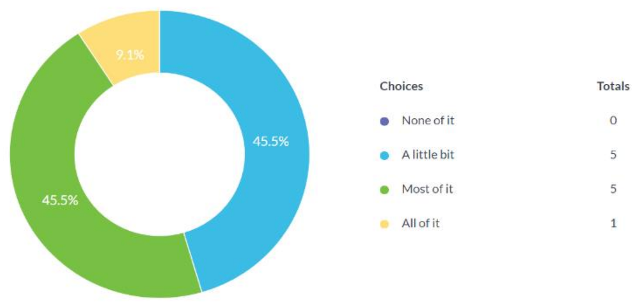


Figure 48

Was the tool nice to look at?

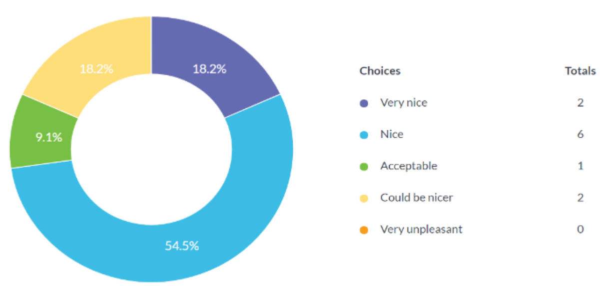


Figure 49

Was the tool nice to use?

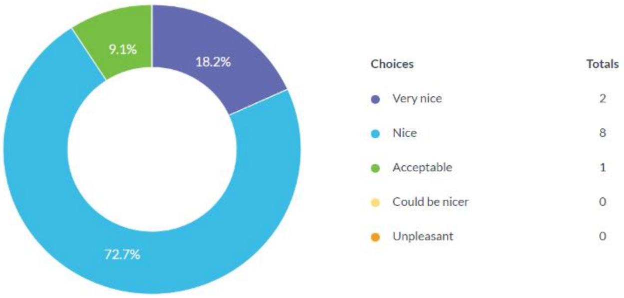


Figure 50

If the tool was not nice to look at or easy to use please expand on the parts that could be improved.

- Space out the field names and text boxes. Perhaps could try using Tenovus Colours? Change the field names to make it a little clearer to the end user what each one does. At the moment its very code like, could it be simplified?
- It was quite easy to use and the menu bars were like other programs which I liked. However, it still felt a little bit like the 'back-end' of a program (which whilst easy to use with the instructions) was a little daunting initially.
- The only slightly complicated or that I can imagine would make it harder to use is getting to grips with the naming of the layers. I don't know how easy this would be to change, but this could perhaps make it an even easier tool to use.

Have the changes to the tool made it easier for you to use?

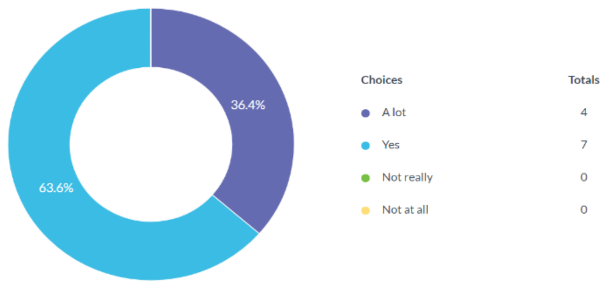


Figure 51

Does the tool fit in with the Tenovus brand?

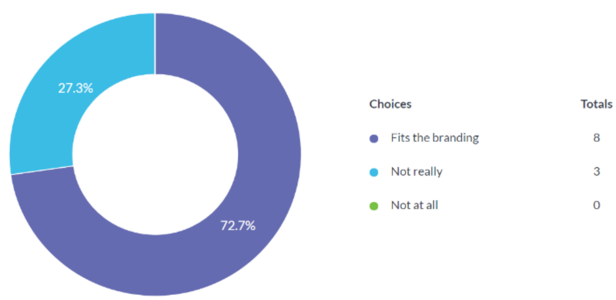


Figure 52

Please describe a task you might use the tool to complete.

- Establish the best place to locate a mobile treatment unit or choir.
- To see what the catchment area of my choirs are.
- Mapping our charity shops and trying to select and prioritise the ones to run research engagement events in.
- To see where our sun cream van visits and where we could potentially visit in the future.
- The tool could be used to help us to understand how much ground we cover (literally), how much 'reach' we actually make with each or a combination of services and potentially where we could go next with a service to increase our coverage. Additionally, a function where we can overlay our services against others that operate in the same area could prove very useful, as we would then be able to understand the different

areas' impact and how much of a difference we could actually make. Using all of this could be useful in a number of conversations, from speaking to potential sponsors, reporting and planning strategically - it could be a real valuable tool for us.

- Tracking where patients are in comparison to outreaches for CSA advice.
- Although I like the tool as an extremely useful piece of software, it is unlikely that I would be using this in my current role.
- To analyse whether our services are best placed to provide the best possible impact.
- Three-fold. In the research team I might use the system to map in the future the impact of our research engagement in specific areas. We might try and see the grants we fund in relation to services offered in the area. Finally, as part of internally evaluation, we would use the multiple functions of the tool to be able to visualise the reach of all the service that we offer across Wales.